Natural Capital Ecosystem Assessment: NE1.6 Change Detection

Scoping object-based change detection for Living England

First published 28 March 2023

Natural England Research Report NECR460

www.gov.uk/natural-england

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Copyright

Catalogue code: NECR460

Report details

Authors

Rory Barber - Ecosystem Analyst, JNCC

Paul Robinson, Earth Observation Applications Manager, JNCC

Natural England Project manager

Anne Stefaniak – Earth Observation Senior Advisor, Natural England

Contractor

JNCC

Keywords

Change detection, object-based, segmentation, accuracy, classification, habitats.

Acknowledgements

The authors would like to thank the following people for generously giving their time to be interviewed for this report: Prof Alexis Comber, Dr Geoffrey Smith, Dr Dan Morton, and Dr Clare Rowland. The authors would also like to thank the DEFRA Library Services Team for their assistance with accessing relevant literature. Finally, the authors wish to thank Lisa Hecker, Izzy Hassall, Lawrence Way, and the Living England Team for quality reviewing the report.

Citation

This report should be cited as:

Barber, R. & Robinson, P. 2023. Scoping Object-based Change Detection for Living England. NECR460, Natural England.

Foreword

The Natural Capital and Ecosystem Assessment (NCEA) is an England-wide programme involving several DEFRA (Department for Environment, Food and Rural Affairs) group organisations. Its purpose is to assess and describe the extent and condition of England's biodiversity, ecosystems, and natural capital assets across our terrestrial, freshwater, and marine environments. To understand the state of our natural capital assets and their ability to provide benefits to people, we need to know how they are changing in extent. Our current ability to accurately measure change in habitat extent is limited, representing a major gap in our understanding of the state of natural capital in England. Improving our ability to measure change in habitat extent will enable the DEFRA to:

1. **Assess progress towards the 'Theme D: Wildlife' goals and targets outlined in the Government's 25 Year Environment Plan** (DEFRA, 2021)**.** The government has committed to establishing a Nature Recovery Network: an increasingly connected network of places that are richer in wildlife and more resilient to climate change. One of the main targets to be assessed is: 'Creating or restoring 500,000 hectares of wildlife-rich habitat outside the protected area network.'
2. **Fulfil the Government's national and international biodiversity reporting requirements.** The D1 indicator in the 25 Year Environment Plan – 'Quantity, quality, and connectivity of habitats' – aims to measure change in extent, condition, connectivity and function of terrestrial and freshwater habitats in England (DEFRA, 2021). Reporting against this indicator will contribute to our reporting commitments for numerous international conventions, such as the Convention on Biological Diversity (CBD).
3. **Evaluate the success of policy interventions**. This includes the roll-out of the Environmental Land Management Schemes (ELMs) which will pay farmers and land managers to deliver environmental benefits on their land.

In support of the NCEA programme, Natural England (NE) is leading the Living England project; a multi-year programme to deliver a satellite-derived national habitat map every two years. The Living England satellite derived habitat map uses an object-based image classification technique known as Object-based Image Analysis (OBIA) which predicts the most likely broad habitat class for parcels of land known as objects. This is done to produce a unit of analysis that more accurately represents real-world features in the landscape, and is therefore more appropriate for habitat classification and the detection of change in habitat extent (Blaschke, 2005). In Living England, objects are produced using a process called image segmentation, which groups pixels of an image together based on the similarity of their spectral information. A multi-resolution image segmentation of Sentinel-2 imagery is used to produce a set of image-objects which are then assigned a *most likely* habitat within a machine learning model using a Random Forest classification algorithm. This algorithm is first trained to identify each habitat class using a set of reference data for each habitat. This reference data is either collected from ground/field data, historical habitat records or desktop surveys of satellite imagery. The output of the classification algorithm is the Living England habitat probability map on a national scale, which is then validated for its accuracy using an independent subset of the reference data. For full details of the methodological approach, please refer to the latest Living England technical guidance (Kilcoyne et al., 2022).

Between each iteration of Living England, the Change Detection project aims to assess changes in the extent of terrestrial habitats. The method to assess change will be developed to produce a highly accurate and reliable dataset for regularly monitoring broad-scale changes in habitats. However, accurately detecting change between classified habitat maps is not as straightforward as simply overlaying one on top of another to identify areas which have ‘changed’ habitat class (Tewkesbury et al., 2015). Real changes need to be separated from changes which may have occurred due to errors and any methodological differences used to produce each map. This requires a clear understanding of the following:

- The methodology used to produce a habitat classification map
- The potential sources of error when producing a habitat classification map
- How these errors accumulate when two habitat classification maps are combined to perform change detection analyses
- Techniques to assess the accuracy of a habitat classification or change detection map
- Appropriate approaches to reduce errors and improve accuracy

An assessment of existing methods and best practice will help develop a standardised change detection methodology for assessing changes between future iterations of Living England maps and aid further assessments of habitat change by NE. The purpose of this review is to:

1. Highlight key issues in existing object-based change detection (OBCD) approaches and identify potential solutions to address these issues.
2. Discuss the assessment of accuracy in OBCD and identify approaches to reduce uncertainty and improve accuracy
3. Make a series of recommendations for performing OBCD to the Living England project team.

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Executive summary

As part of the NCEA programme, JNCC is supporting NE by conducting a series of expert interviews and a literature review to identify best practices to assess changes in broad-scale habitat distribution and extent from Earth Observation (EO) derived OBIA datasets. The key source of issues identified by the review with respect to OBCD was the spatial framework used to detect change. A spatial framework is required in habitat change detection to assess whether there have been any spatial changes in a habitat's size and shape, or whether there have been any thematic changes, such as the conversion from one habitat class to another. The spatial framework used by Living England are the image-objects produced during the image segmentation process. A single spatial framework using image-objects from one Living England iteration can be maintained over the entire change detection period, facilitating the ability to detect if an object has changed habitat class or not. However, keeping the same spatial framework implies that the image-objects have not changed size or shape, meaning it is difficult to detect changes in the distribution and extent of habitats. Alternatively, the spatial framework can be changed, using a different set of image-objects produced for each iteration of Living England. This supports the identification of changes in object boundaries, enabling changes in habitat extent and distribution to be detected. However, the image segmentation process used to produce each set of image-objects is not 100% accurate and is reliant on the quality and acquisition conditions of the input imagery. This means that changes in the boundaries of image-objects may not always reflect changes in the boundaries of real-objects on the ground. When comparing the two sets of image-objects for change detection, millions of tiny sliver polygons will be produced from these small boundary changes. Separating the real habitat changes from false changes which have occurred due to possible errors during the image segmentation becomes practically challenging.

Deciding whether to maintain or change the spatial framework over the change detection period is a key consideration for Living England. Numerous solutions have been proposed to address the issues associated with each approach, each with their own advantages and disadvantages. For example, sliver polygons below a certain size can be assumed to be produced from errors and filtered out, but this process risks losing important information about small-scale changes in habitat boundaries. Another option is to use a single set of image-objects produced by combining imagery from multiple iterations of Living England and classifying changed objects directly. However, this makes it more challenging to detect the type of change that is occurring because only one classified product is generated, as opposed to comparing two classified products at different time periods. Alternatively, a spatial framework could be derived from digital cartography, such as Ordnance Survey Maps, which are regularly surveyed and updated. Although a spatial framework derived in this way is more likely to reflect real boundaries, there remain issues over licencing and the identification of habitat boundaries in upland areas which are less regularly mapped. Whichever approach Living England chooses, there will be limitations that need to be considered and carefully communicated to end users. The most practical solution is to retain all the raw input imagery used in the production of each Living England iteration to enable flexibility over the long term. Remote sensing methodologies advance

rapidly; retaining the raw imagery enables new advances and approaches to spatial frameworks to be trialled and tested using historical imagery to determine their potential suitability for change detection. Any one spatial framework can then be chosen from a Living England iteration and used to re-analyse historical imagery and detect change using a consistent spatial framework.

Habitat classification maps need to be assessed for their accuracy by comparing them to reference data to demonstrate their integrity and real-world applicability. This is regularly done for individual maps, Living England Phase IV had an 88% agreement between the classification map and reference data (collected from the field) (Kilcoyne et al., 2022). Although this seems high, this overall figure hides important variation between habitats and geographic location, which must be communicated to the end user. For example, Acid, Calcareous and Neutral Grassland has a much lower accuracy in the south-west of England due to difficulties distinguishing it from the Improved Grassland habitat (Kilcoyne et al., 2022). Due to this variation in accuracy, it is not possible to simply combine two habitat classification map accuracies together to get an accuracy figure when performing change detection between two maps. The change map produced during change detection needs to be independently assessed for its accuracy. This is challenging because the amount of reference data required to assess the accuracy of a change map is significantly higher than the amount required to assess the accuracy of a single habitat classification map. Living England contains 17 habitat classes, each of which requires a set of reference data to validate how accurate the classification of that habitat is. In a simplified map of habitat change, there will be at least 272 (17^2 - 17) possible changes that could have occurred between habitat classes, and 17 additional instances where a habitat class doesn't change, each of which requires a set of reference data to validate the predicted change or no change. This is because each habitat can either stay the same or change to one of the other 16 habitats. This excludes additional 'transitional' habitats, which don't fit into the 17 main habitat classes, but represent a transitional state as a habitat changes class from one to another. Clearly this is a significant increase in the resources required for gathering reference data and validating a habitat change map.

Reducing the amount of reference data that needs to be collected is the most practical solution to the problem of accuracy assessment. This can be done via several approaches, but fundamentally they all rely on having clear objectives for the habitat change map. There is a clear need to understand what types of habitat changes the map will be able to identify based on its technical inputs, who the end user will be and what the map will be used for. For example, the user requirements of a local nature reserve manager will be quite different to a national policy advisor. Once these objectives are established, a sampling strategy to collect reference data can be developed which prioritises certain transitions or the accuracy of certain transitions over others (Olofsson et al., 2014). For example, the transition between Urban and Water habitats may not be considered relevant to the objectives of the map and so reference data isn't collected for it. Unlikely transitions could also be removed, such as between Coastal Sand Dunes and Arable habitats. Alternatively, habitat classes could be combined if necessary, such as Coniferous and Broadleaved Woodland, detecting change between broader categories of habitat such as Grassland and Woodland. Finally, transitions between habitat classes with low accuracy scores could be removed if there is insufficient confidence in the ability to detect change between them. The sampling strategy used to collect reference data is crucial to ensure that changes can be appropriately assessed for their accuracy; further details on designing such a strategy can be found in (Olofsson et al., 2014).

The review concludes by making the following set of recommendations for Living England to consider in their development of a change detection methodology:

1. Clearly define the objectives of the change detection product based on the scale, thematic and end-user requirements of the outputs.
2. Maintain the spatial framework from one of the input maps and use this to analyse imagery from both time points to avoid the production of sliver polygons. Further investigate approaches to detect changes in habitat boundaries when maintaining the spatial framework.
3. Retain all the raw input imagery to enable flexibility when performing change detection. Historical imagery can then be reanalysed using a single spatial framework taken from any iteration of Living England.
4. Reduce the number of transition classes being detected by the change map if necessary. This reduces the amount of reference data that needs to be collected to assess the accuracy of the change detection product.
5. Design a robust sampling strategy to collect reference data and independently validate transitions identified in the change detection product.
6. When communicating uncertainty, present more than just overall accuracy figures for the change detection product.
7. Consider the use of alternative change detection and accuracy assessment approaches if object-based change detection between two habitat maps is not well suited to the user objectives.

Contents

Table of figures 10

1. Background 11

   1.1 - The Living England project 11

   1.2 - Object-Based Change Detection (OBCD) 12

2. Aims 14

3. Methodology 16

4. Results and Discussion 18

   4.1 - Image segmentation and spatial frameworks 18

   4.2 - Maintaining v changing the spatial framework 20

   4.3 - Possible options to address spatial framework issues 22

   4.4 - Accuracy and uncertainty in change detection 27

   4.5 - Approaches to support accuracy assessment in change detection 33

5 - Recommendations 40

6 - References 43

Annex 1: Literature Review process 47

List of abbreviations 50

Table of figures

Figure 1: An overview of different change detection methods and their respective advantages and disadvantages. Most can be applied to either pixels or objects. Source: Tewkesbury et al. (2015). 13

Figure 2: Separate examples of (a) over, and (b) under segmentation when generating land parcel objects from imagery. In (a) the group of five segments in the centre should have been merged into a single field. In (b) the large segment in the centre should have been subdivided into at least two separate fields. Source: Smith and Morton (2010) under CC BY-NC-ND 4.0 LICENCE....19

Figure 3: A hypothetical error matrix detailing the calculation of Overall, Users and Producers Accuracy for three classes: Forest, Urban and Water. The F:F, W:W and U:U boxes represent the number of correctly classified objects. Source: Story and Congalton (1986), reprinted with permission from the American Society for Photogrammetry & Remote Sensing, asprs.org. 28

Figure 4: Equations used to compare pairs of maps to assess differences due to change and those due to errors; also to identify changes hidden by errors. Source: Fuller et al. (2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier. 30

Figure 5: Assessing likely differences due to change and those due to errors between Land Cover Map 1990 and 2000. Source: Fuller et al. (2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier. 31

Figure 6: The proportional area of expected change which can be measured with 75% reliability when using pairs of maps with varying class numbers and map accuracies. As the proportional area of expected change declines, the higher the accuracy of the maps required in order to reliably detect the change. Source: Fuller et al .(2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier. 32

Figure 7: Mean overall accuracy and mean proportion of total area classified as ‘change’ in high, medium, and low sensitivity change detection maps created by object-based analysis of multi-temporal aerial and LiDAR imagery (*n* = 13). Error bars represent standard deviation. Source: Lightfoot et al. (2020) 36

Figure 8: An example of the spatial variation in producer accuracy for the 'Grazing Land' class. Crosses indicate locations where Grazing Land was recorded in the reference data, whilst circles indicate locations where Grazing Land was recorded in the classification. Source: Comber (2013) 38

1. Background

1.1 - The Living England project

The Natural Capital and Ecosystem Assessment (NCEA) is an England-wide programme involving several Defra group organisations. Its purpose is to assess and describe the extent and condition of England's biodiversity, ecosystems, and natural capital assets across our terrestrial, freshwater, and marine environments. To understand the state of our natural capital assets and their ability to provide benefits to people, we need to know how they are changing in extent. However, our current ability to accurately measure change in the extent of habitats is limited, representing a major gap in our understanding of the state of natural capital in England. Improving our ability to measure change in habitat extent is crucial for assessing the impact of policy interventions, such as the new Environmental Land Management schemes, and reporting against the D1 indicator – Quantity, quality, and connectivity of habitats – in the Government's 25 Year Environmental Plan (DEFRA, 2021). Further to this, ambitious targets have been set within the Environment Act to restore or create 500,000 hectares of wildlife-rich habitat (DEFRA, 2022).Ongoing data provision will be vital to enable assessment of change to determine if progress is being made towards that target. Creation of new habitats will mean a change of habitat class in subsequent habitat maps; however restoration of habitats may not represent a change of habitat class but a change of condition within one class. Both creation and restoration will need to be assessed, and the methods discussed here may support that assessment.

In support of the NCEA programme, NE is leading the Living England project; a multi-year programme to deliver a satellite-derived national habitat layer, which assesses the extent and distribution of England's terrestrial habitats and natural capital assets. The project is also exploring potential approaches to assess change in habitat extent using iterations of the Living England dataset. The Living England project is currently in its fifth phase, building on the methodological developments and recommendations from the first three phases of the project. Phase I (completed March 2019), developed and tested a process to apply the Living Maps methodology at a national scale (Kilcoyne et al., 2020). Phase II (completed March 2020) implemented key improvements to build on the national-scale method, including the introduction of biogeographic zones (Kilcoyne et al., 2020). Phase III focused on establishing and deploying standardised methods for collecting field training data for the model and further developing the classification process (including image generation, segmentation and validation) (Kilcoyne et al., 2021). Finally, Phase IV updated the segmentation process, and improved elements of the classification method for the categories of: built-up and gardens, coastal saltmarsh, water, bare ground, and arable and horticulture. It also introduced a new approach to deal with cloud masking and is the first version of Living England to be published under an Open Government Licence (Kilcoyne et al., 2022).

The Living England project uses an object-based machine learning approach to image classification, developed under the Defra Living Maps project (SD1705 – Kilcoyne et al., 2017). Broadly, the method first clusters homogeneous areas of an image into segments based on spectral, spatial and other contextual information, using a process known as image segmentation. Reference field samples are collected for each habitat class and aligned with the segments produced in the first step. Image statistics from these sample segments and ancillary datasets (e.g. climate and topographical information) are then extracted from each of the input imagery layers and used to train a Random Forest classifier (a machine learning algorithm). The classifier then assigns each image segment to *the most likely* habitat class based on this training data, identifying the probability that the segment belongs to the first and second most likely class. Some habitat classes are defined using pre-existing datasets via a vector-based classification approach, including arable and horticultural from the Crop Map of England (CROME - Rural Payments Agency), Saltmarsh Extent and Zonation (Environment Agency) and Ordnance Survey Vectormap for built-up areas and gardens. Finally, the outputs of the classifiers are merged and the model-based classification is assessed for its accuracy using the validation subset of reference samples. For full details of the methodological approach, please refer to the latest Living England technical guidance (Kilcoyne et al., 2022).

1.2 - Object-Based Change Detection (OBCD)

Object-based approaches to land cover classification, such as the Living England methodology, have become frequently used as an alternative to traditional pixel-based classification approaches, which classify individual pixels instead of segments (Blaschke et al., 2014). Although pixels are the fundamental unit of any image, advances in the resolution of satellite-derived imagery mean that pixels are often now smaller than objects of interest and no longer reflect meaningful features in the landscape - a single field or patch of woodland may contain thousands of pixels. Contrarily, object-based image analysis (OBIA) groups spectrally similar pixels into more ‘meaningful’ image-objects which aim to represent real-world features in the landscape (Smith and Morton, 2010). These image-objects provide a more relevant unit of analysis for performing image classification and change detection analysis with high resolution imagery, such as assessing change in habitat extent across landscapes. Raw pixels obtained from satellite-imagery contain limited comparable information, namely measures of tone and radiance. By contrast, the use of image-objects enables the comparison of additional properties such as size, shape, pattern and texture; all of which provide important contextual data which can improve the image classification process (Tewkesbury et al., 2015). Furthermore, pixel-based classification and change detection approaches are prone to the production of spurious pixels - what is known as the ‘salt-and-pepper’ effect (Liu and Xia, 2010). This occurs due to the sole use of spectral information during the image classification process, which can result in high variability in the spectral data values used to define a habitat class. If this spectral variability overlaps with the spectral range of another class, then this can result in the incorrect classification of a pixel and the incorrect detection of change in a pixel. To overcome this issue and create a more meaningful unit of analysis there has been a rise in the number of studies using object-based change detection (OBCD) methodologies since the turn of the century (Blaschke, 2005; Blaschke et al., 2014). Although pixel-based and object-based approaches differ in their unit of analysis, it should be noted that many classification techniques and change detection analyses can be applied to either pixels or objects (Figure 1 - Tewkesbury et al., 2015).

	Description	Advantages	Limitations
Layer arithmetic	Image radiance or derivative features are numerically compared to identify change.	Can be simple to implement.	Usually gives little insight into the type of change.
Post-classification change	The comparison of multiple maps to identify class transitions.	Produces a labelled change map. Prior radiometric calibration may not be required.	Errors in any of the input maps are directly translated to the change map.
Direct classification	A multi-temporal data stack is classified directly identifying both static and dynamic land covers.	Only one classification stage is required. Provides an effective framework to mine a complicated time series. Produces labelled change map.	Classification training datasets can be difficult to construct, especially for a time series of images.
Transformation	A mathematical transformation to highlight variance between images.	Provides an elegant way to handle high dimensional data.	There is no defined thematic meaning to the results. Change may be difficult to locate and interpret.
CVA	The computation of difference vectors between analysis units giving both the magnitude and direction of change.	Gives insight into the type of change occurring.	In its raw form the change direction and magnitude may be ambiguous.
Hybrid change detection	The use of multiple comparison methods within a workflow. The most commonly used strategy is a combination of layer arithmetic to identify change and direct classification to label it.	Training data does not have to be collected over radiometrically stable areas.	No specific limitations.

Figure 1: An overview of different change detection methods and their respective advantages and disadvantages. Most can be applied to either pixels or objects. Source: Tewkesbury et al. (2015).

2. Aims

The Living England project aims to produce a national habitat probability map every two years to regularly assess the extent of terrestrial habitats and natural capital assets. Between each iteration of the national habitat layer, Living England wants to assess changes in habitat extent and distribution. The Change Detection work package thus aims to develop the use of the Living England data from multiple years to assess change in habitat extent. The method to assess change will be developed to produce a highly accurate and reliable dataset for regularly monitoring broad scale changes in habitats. This aims to fulfil policy and reporting requirements and inform programmes such as the Environmental Land Management (ELM) scheme. There are a wide range of change detection methods available with their respective advantages and disadvantages (Figure 1). Accurately detecting change between classified habitat maps is not as straightforward as simply overlaying one on top of another to identify areas which have 'changed' habitat class (Tewkesbury et al., 2015). Real changes need to be separated from changes which may have occurred due to technical and semantic differences in the methodological approach used to produce each map. This requires a clear understanding of the sources of classification error, how these propagate in object-based change detection and appropriate approaches to mitigate and reduce these errors.

As part of the NCEA programme, JNCC is supporting NE by conducting a review of best practices to assess changes in broad-scale habitat distribution and extent from Earth Observation (EO) derived OBIA datasets. An assessment of existing methods and best practice will help develop a standardised change detection methodology for assessing changes between future iterations of Living England maps and aid further assessments of habitat change by NE. Further work is being done by JNCC using Living England alongside time series of satellite data, to highlight changes in condition of habitat parcels, even when that does not generate an overall change in class.

This review therefore aims to highlight key issues in existing approaches, identify potential solutions and make recommendations to the Living England project team. In addition to identifying key issues and potential solutions associated with OBCD, Natural England identified two priority issues that this report will address in detail. These were:

1. Detecting change to habitat distribution and extent using image segmentation. The Living England habitat probability map utilises an OBIA segmentation process with Sentinel-2 satellite imagery using Trimble eCognition software (Kilcoyne et al., 2022). The segmentation is based on the grouping of similar spectral signatures in addition to shape, geometry, and additional parameters. The segmentation process will therefore have a key influence on the extent and quantification of the habitats. Thus, when assessing change in habitat extent from the Living England outputs, it is essential to consider whether the segmentation (i.e. the spatial framework) should be repeated for each iteration to detect change in boundaries as well as habitat cover, or if the original segmentation should be maintained across iterations. The approach chosen will have implications for both segment and sub-segment change.

2. Assessment of uncertainty associated with using classified probability habitat maps for detecting and measuring change. The Living England maps provide national coverage of the most likely habitat within each segment. When comparing two classified maps to assess change, otherwise known as post-classification change detection, it is likely there will be additional uncertainty and areas of reduced accuracy. Recommendations for improving accuracy and reducing uncertainty with post-classification change detection should be discussed.

3. Methodology

To identify key issues and possible solutions with respect to OBCD, the review used two complementary approaches: an in-depth literature review, and a series of interviews with experts in the field of land cover mapping and change detection. An overview of each of these approaches is given below, with further details of the literature review process provided in Annex 1.

Due to the large volume of literature published on the topic of OBCD (939 articles are returned from just the keyword ‘Object-Based Change Detection’ on the academic database Scopus as of 02/02/2022), and the time constraint associated with the project (5 weeks FTE), a comprehensive systematic review approach was not chosen. Instead, the literature review focused on the use of known, relevant and recent review articles on the subject matter identified by NE and JNCC; literature identified and recommended from the expert interviews; and a time-limited literature search of the Scopus database for further literature. In total, 100 articles and reports from academic and public-body sources were identified as relevant for the review, with 56 being published under Open Access licences. Due to time constraints, 35 articles and reports were subsequently read by the authors to identify relevant key issues and possible solutions related to OBCD. Having prioritised much longer lists of articles on the topic it was felt that these articles covered the significant developments in the area and would provide the best review given the time available. For full methodological details of the literature review, including search string development, relevance screening criteria and articles consulted, please refer to Annex 1.

In addition to published literature, three interviews were arranged with known experts in the field of land cover mapping and change detection. These included people with extensive experience in the production of national and international land cover maps as well as academics who have published widely about change detection using EO. Interviewees and dates for each of these interviews are outlined below:

1. **Dr Geoffrey Smith (Specto Natura) - 25/11/2021**

   Dr Smith has over 25 years’ experience in the field of EO-based applications including leading the UK Centre for Ecology and Hydrology (UKCEH) EO team creating national Land Cover Maps. He has also been part of the consortium creating the CORINE Land Cover Map for the UK as well as being a member of the EIONET Action Group on Land monitoring in Europe who work to harmonise data from European and national data sources for land cover.

2. **Dr Dan Morton and Dr Clare Rowland (UKCEH) – 06/12/2021**

   Dr Morton and Dr Rowland are Earth observation scientists working for UKCEH. Both have many years’ experience in the production of national Land Cover Maps, the development of methods for producing national level products from EO and the production of change products from the Land Cover Map production process.

3. **Prof Alexis Comber (University of Leeds) – 28/01/2022**

   Prof Comber is Professor of Spatial Data Analytics at Leeds Institute for Data Analytics (LIDA) the University of Leeds. He has many years experience in the analysis of uncertainty in spatial data and the validation of EO-derived data products.

Each interview lasted an hour and started with a pre-defined set of open-ended discussion points to guide the conversation. It is worth noting that as more literature was reviewed on the subject, later interviews incorporated more technical questions focused on specific components of change detection.

Initial Questions

- How would you address the issue of spatial frameworks when using OBIA for change detection? Should it be updated or remain constant?
- What are the pitfalls of utilising 2 or more iterations of OBIA products to identify change?
- How would you deal with uncertainty in addressing change that way?
- The classification is produced primarily by a Random Forest algorithm, can that be a benefit or drawback in analysing uncertainty?
- What key research are you aware of in this area?

4. Results and Discussion

The stakeholder interviews and literature review identified a range of issues with OBCD and highlighted numerous proposed approaches to improve the accuracy of change detection. This section is divided into several sub-sections. First, an overview is given of the image segmentation process and the use of image-objects as a spatial framework for assessing change. Two sub-sections follow, one for each of the two main issues identified by NE – detecting habitat extent change using OBIA and segmentation methods, and assessing uncertainty when using that approach. General issues related to OBCD are discussed throughout.

4.1 - Image segmentation and spatial frameworks

4.1.1 - What is a spatial framework?

Change detection requires a spatial framework to assess thematic and spatial changes that may have occurred between classification maps. That framework can be a regular grid, such as the pixels of an image, or a set of polygons, such as the image-objects created in the segmentation process. The literature and interviewees refer to changing or maintaining the spatial framework; consequently, the following discussion refers to the spatial framework used in OBCD, which is generally referring to a set of polygons. In the Living England context, the image-objects derived from the image segmentation process form the spatial framework over which change will be assessed. To understand the issues associated with changing or maintaining this spatial framework, it is therefore important to provide an overview of how image segmentation is performed and what an image-object represents. Full details of the Living England segmentation process can be found in the Phase IV technical guidance (Kilcoyne et al., 2022).

4.1.2 - Approaches to image segmentation

There are two main approaches to image segmentation: top-down and bottom-up. Top-down approaches (such as multi-scale wavelet decomposition) start with a single region of pixels and progressively break it up into smaller objects until a terminal condition is met. Conversely, bottom-up approaches (such as region growing, fuzzy clustering and thresholding) start with individual pixels and group them together until a condition is met (Shepherd et al., 2019). The segmentation algorithm uses a set of user-defined parameters such as scale, shape, and compactness, which determine how the algorithm divides the image up. Although studies have attempted to find ‘optimal’ parameters for segmentation algorithms, segmentation is often applied in specific contexts and so parameterisation is difficult to generalise (Shepherd et al., 2019; Smith and Morton, 2010). Identifying input image bands and parameter values that produce image-objects which closely align with real-objects on the ground is therefore often a process of trial and error by the user, although studies have attempted to quantify the accuracy of a given segmentation (Liu and Xia, 2010). Image segmentation thus produces two main types of error: over-segmentation and under-segmentation. Over-segmentation occurs when the image-objects are smaller than real-objects on the ground, and under-segmentation is the opposite, when the image-objects are larger than real-objects (Figure 2). In the context of classification, under-segmentation can result in different classes merging together, so users normally prefer to have a slight over-segmentation to retain classification accuracy (Liu and Xia, 2010).

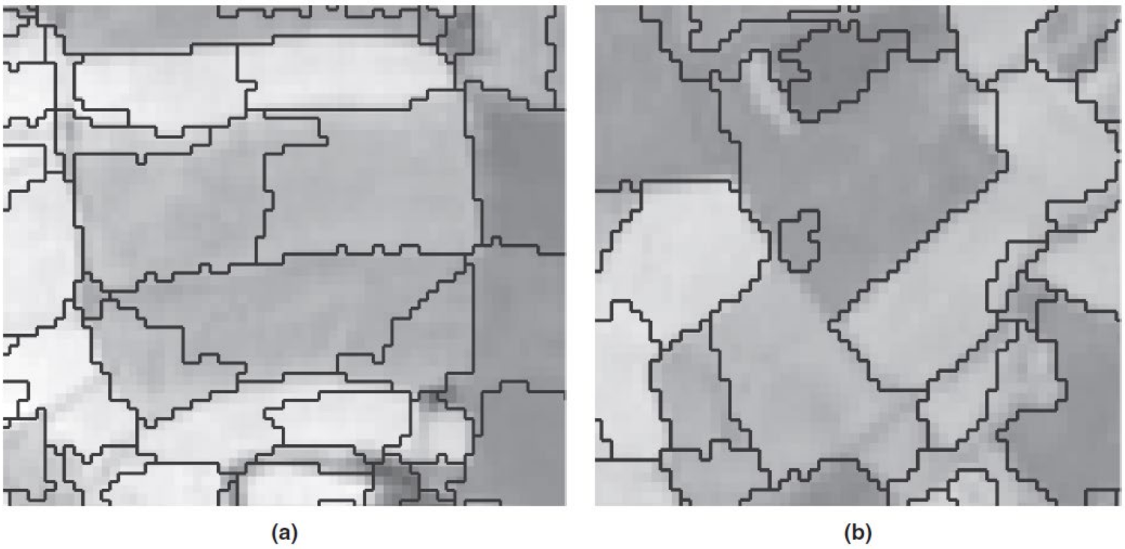

Figure 2: Separate examples of (a) over, and (b) under segmentation when generating land parcel objects from imagery. In (a) the group of five segments in the centre should have been merged into a single field. In (b) the large segment in the centre should have been subdivided into at least two separate fields. Source: Smith and Morton (2010) under CC BY-NC-ND 4.0 LICENCE.

4.1.3 - Limitations of using image-objects as a spatial framework

The key point about image-objects is that they are *directly derived* from a set of input imagery. Variations in the conditions under which imagery are collected, such as time of day, sensor properties, seasonal changes in phenological responses of vegetation and extreme events can affect the input imagery and thus the outputs of image segmentation, even if land cover remains stable (Tewkesbury et al., 2015). Although Living England uses multiple images from different seasons in its segmentation process, an extreme event or a prolonged change in weather may affect the phenological responses of vegetation and alter their spectral signature in the imagery. Consequently, image-objects are only relevant to the specific timeframe over which the imagery was collected, rather than being relevant *through* time and between iterations (Comber and Wulder, 2019; Smith and Morton, 2010). This is important to remember in the context of change detection and whether to maintain or change the segmentation for each iteration of Living England.

4.2 - Maintaining v changing the spatial framework

4.2.1 - Maintaining the spatial framework

Overlaying the image-objects produced by the segmentation of one set of imagery onto another set of imagery creates a simple unit of analysis for post-classification change detection image-object overlay. The spatial framework is maintained through time and information such as tone and texture from the two sets of imagery can subsequently be compared quite easily (Tewkesbury et al., 2015). However, if image-objects are only relevant at a given point in time, then maintaining the spatial framework limits the ability to identify several meaningful changes in habitat extent. First, changes in the boundaries of objects, in this case habitats, become undetectable. Using the same image-objects implicitly assumes that the boundaries of those objects have not changed over the time-period of comparison. Second, as a result of not identifying boundary changes, sub-object changes in class may not be detected unless the change is large enough for the entire object to be classified as a different class (Hall and Hay, 2003). Changes in habitat class of an entire object may be detected, such as the conversion of a woodland into farmland, but smaller changes occurring within objects will not be detected, such as the planting of a small patch of woodland on farmland. This has important implications for understanding habitat dynamics and estimating the total area of habitats in the UK, where 71% of land is used for agriculture and most habitat change is small in scale and occurs at the margins (Hayhow et al., 2019; Lawton et al., 2010). Finally, if assessing change over longer periods of time, the original boundaries of the image-objects used as the spatial framework may become less relevant and accurate, as the landscape undergoes further change. For example, the boundaries of image-objects derived from 2022 imagery are unlikely to still be accurate in 2052, meaning that change detection is unlikely to provide results that are relevant to existing policy and programmes.

4.2.2 - Changing the spatial framework

The alternative to maintaining a spatial framework for change detection is to generate new image-objects for each iteration of Living England, creating independent spatial frameworks. This enables the spatial framework to remain relevant to each iteration of Living England and enables the comparison of additional object information, such as size, shape, pattern and association (Tewkesbury et al., 2015). Importantly, sub-object changes may become detectable (if the changes are greater than the spatial resolution of the input imagery) and changes in habitat boundaries can be identified, facilitating a greater understanding of habitat dynamics. However, linking these independent spatial frameworks through time is challenging. As previously identified, the outputs of image segmentation are dependent on the input imagery, which can be affected by image acquisition conditions, seasonal phenology, and other atmospheric conditions. Although the Living England method uses multi-seasonal image composites taken throughout the year, the effects of extreme weather events can last several years. Consequently, the phenological response of the vegetation may change, altering the spectral signature detected by the satellite imagery and impacting the outputs of any subsequent segmentation process. This could lead to pixels being incorrectly clustered together during the segmentation. Furthermore, random elements in the operation of the segmentation algorithm itself, such as the selection of seed points used to define the starting location of a region-growing segmentation algorithm, can generate different image-objects from the same input imagery, although this is not always the case. Therefore, even if real-objects have not changed, the image-objects themselves might have changed, producing inconsistent object-boundaries. When comparing inconsistent objects through time, small sliver polygons are produced due to the differences that occur between object boundaries at two time periods. Importantly, these differences could reflect:

- Real changes in habitat boundaries on the ground
- Incorrect changes due to misalignment between the two sets of imagery
- Incorrect changes due to differences in the spectral reflectance captured in the segmentation input imagery
- Incorrect changes that occur from operational variation in the segmentation algorithm (Smith and Morton, 2010).

Importantly, this raises the possibility of falsely identifying areas of change. Handling these sliver polygons is practically challenging; all our interviewees agreed that this approach would be considered the most difficult to apply at a national scale. Given Phase IV of Living England produced nearly 7 million polygons, it is likely that millions of tiny sliver polygons would be produced during an image-object overlay due to the issues described above (Kilcoyne et al., 2022).

4.3 - Possible options to address spatial framework issues

Our stakeholder interviews and review of the literature has identified numerous possible approaches that could be taken to address some of the issues associated with changing or maintaining the spatial framework for OBCD. The sub-sections below are divided into those which tackle issues associated with maintaining and changing the spatial framework.

4.3.1 – Approaches for maintaining the spatial framework – multi-temporal segmentation

As discussed, one issue associated with maintaining a spatial framework for change detection is that objects are not relevant over the entire change detection time-period. One approach to address this is to perform a single image segmentation using multiple images collected over the entire change detection period, otherwise known as multi-temporal image segmentation. Living England already uses this approach during its segmentation process – it uses multiple images collected throughout the year to minimise the effects of seasonal changes in vegetation phenological responses. This approach would combine the input imagery used in each iteration of the Living England product and perform a single segmentation over the entire set of imagery. In doing so, the resulting image-objects would honour persistent object boundaries and no sliver polygons would be produced during change detection because only a single segmentation is performed (Tewkesbury et al., 2015). Pre-classification change detection techniques can then be performed on the entire set of input imagery, using a single set of image-objects. Image statistics are then calculated for each image within the set of imagery and used to identify areas which change. This produces a labelled change map, identifying areas where the habitat class has remained stable and areas where it has changed. This change map would then need further classification, either using a classifier or a set of knowledge-based rules, to determine the type of habitat change (Lightfoot et al., 2020).

Several studies in forest habitats have used this approach to identify changes using long-term time-series of imagery (Bontemps et al., 2008; Desclée et al., 2006). For example, Desclée et al. (2006) extracted reflectance band values for each image-object from each of the image dates used in a multi-temporal segmentation. This enabled them to compare spectro-temporal differences in image-object statistics, such as mean and standard deviation values for each spectral band. Using these values, they could then perform a statistical process, multivariate iterative trimming, to identify objects that had statistically changed during the time-series. They performed direct classification on the multi-temporal stack of images and produced a labelled change map, identifying image-objects that remained stable and others which changed class.

Although this approach resolves the issue of objects becoming irrelevant through time and avoids the production of sliver polygons, it fails to address other important issues. First, this approach does not enable the use of post-classification change detection, meaning that although changes in object size and shape can be detected, they cannot be easily compared through time (Tewkesbury et al., 2015). Furthermore, small changes to object boundaries may be lost during the segmentation process, limiting the ability to detect small scale changes in habitat. Second, the classification techniques that can be applied only show the magnitude and direction of change rather than the specific nature of any transitions (from-to), although knowledge-based rules can then be used in order to determine the nature of change (Lightfoot et al., 2020). Applications of this approach have therefore been limited to relatively constrained land cover changes, such as detecting deforestation and urban expansion (Bontemps et al., 2008; Liu et al., 2021). Finally, the construction of robust reference data – used for training and validating the classifier - over the entire change detection period becomes challenging, particularly for areas which have changed. These data would have to be representative for the entire change detection period given that only a single classification stage is performed. This may have consequences for the accuracy of the overall classification since accuracy measures are calculated from reference datasets. Consequently, multi-temporal segmentation approaches to change detection have been the focus of more recent developments in the use of deep learning to perform unsupervised classification without training data. However, to date these studies have been limited to small test sites, typically in urban areas, examining a limited number of potential transitions and deploying very high-resolution imagery of around 1 m (Liu et al., 2021; Song et al., 2020). Although there have been limited studies comparing the results of pre- and post-classification change detection methods, those to date suggest that post-classification methods achieve similar or higher overall accuracies than pre-classification methods (Lightfoot et al., 2020).

4.3.2 - Approaches for changing the spatial framework – Removing slivers

As previously described, a significant problem with identifying change between two OBIA products with frameworks produced by segmentation is the issue of sliver polygons produced by the overlay operation. The principal approach adopted by some studies to deal with this is to perform a series of post-processing techniques to merge and filter sliver polygons based on a set of predetermined rules (Tewkesbury et al., 2015). Objects can be smoothed to eliminate long-thin slivers arising from differing image-object boundaries, whilst small objects can be filtered out entirely based on a predetermined area threshold (Zhang et al., 2017). Sliver polygons below the area threshold are assumed to be arising from inconsistencies in the segmentations, whereas those above the threshold are assumed to reflect actual changes occurring on the ground. Boundary polygons are then smoothed or merged with neighbouring objects, but this needs to consider the similarity of their spectral reflectance values.

The key challenge with this approach is determining suitable size thresholds for the specific context and validating the merging and filtering processes (Shepherd et al., 2019). This is important to ensure that polygons are merged accurately and that sliver polygons are correctly removed. Studies often use the Minimum Mapping Unit (MMU) and a Minimum Feature Width to assist this process; however, this means that polygons below a certain size are automatically removed, which could reflect real changes occurring

(Tewkesbury et al., 2015). Living England has a MMU of three pixels or 300 m^2 (Kilcoyne et al., 2022); removing objects below this size could result in a substantial loss of information about the UK's habitat boundaries and extent. The results of any such filtering and merging process would then require further validation to ensure that each sliver polygon has been correctly merged or eliminated from the change map. Phase IV of Living England produced approximately 6.9 million polygons, a large number to first filter/merge and then validate.

4.3.3 - An alternative approach to handling sliver polygons

Another alternative to handle sliver polygons would be to impose a fine grid over the map products that represent the start and end of the time-period for which change is being assessed. The grid will be consistent across time and each cell would be allocated to a class based on the Living England data at each time-period, enabling each cell to be easily compared across time. The resolution of the grid would need to be high so it is relevant to the input data. In this case the spatial frameworks are generated from 10 m resolution data, so a resolution of 10 - 20 m would likely be appropriate. However, this would require analysis to find a solution that works best. This approach effectively converts the analysis into a raster process and could be done by converting each map to a consistent raster grid. However, retaining the data in a vector format would enable the retention of attribute data relating to probabilities from the random forest classifier or about the proportion of each grid cell covered by different categories when it is compared to the Living England data. This approach does not actually solve the problem of sliver polygons, it simply gives a mechanism to deal with the data in a consistent way. Probability values from the original classifications step could then be generated at the same resolution and could be used in the assessment of uncertainty in change analyses.

4.3.4 - An alternative to a segmentation-based spatial framework?

An alternative solution proposed to address the issues associated with segmentation-based spatial frameworks and the relevance of image-objects through time is to assess change using a spatial framework derived from digital cartography. Smith and Morton (2010) argue that in countries with well-surveyed and regularly updated cartographic maps of the landscape, such as the UK, it is practical to use this as the spatial framework for OBIA. Most landscape features, such as roads and field boundaries, have remained relatively stable in the UK over several decades. Therefore, instead of trying to produce the perfect set of image-objects from the segmentation process to reflect these boundaries, digital cartography may be better suited to use as a consistent spatial framework (Smith and Morton, 2010). This is the approach adopted by UKCEH in the production of Land Cover Map, where a pixel-based classification is first performed, before summarising the pixel-based classification based on objects derived from digital cartography (Morton and Rowland, 2015). Performing a pixel-based classification facilitates the detection of sub-object changes whilst still using a spatial framework to summarise pixel statistics and compare change through objects. Furthermore, any inconsistencies identified between EO data and the cartographic framework could support the detection of real changes that have occurred on the ground.

4.3.5 - Limitations of a digital cartographic approach

There are several limitations with using a spatial framework derived from digital cartography for habitat change detection. First, it should be noted that although UKCEH uses a digital cartographic spatial framework for Land Cover Map, it has not yet developed a robust and consistent method for change detection using a digital cartographic framework. Second, a digital cartographic framework would likely be derived from Ordnance Survey (OS) MasterMap, which is a licence restricted product. If Living England is derived from it, then it may also become licence restricted. This approach would require generalizing detailed cartography to create objects that are appropriate for the desired product. This would be particularly difficult in urban areas as these are likely to be far too detailed. Such a generalisation process could be both costly and need the collaboration of OS, meaning that the production process would have to be agreed with them and licence access to the outputs negotiated. Third, upland areas are not as regularly mapped by the OS and generally have fewer boundaries delineating objects across the landscape. This means that objects are more likely to contain mixed classes, which is more difficult to accurately assess in change detection, and therefore an internal segmentation of upland objects may be required, reintroducing issues previously described. Fourth, updates to the cartographic framework would be needed to maintain its accuracy. This would re-introduce sliver polygons, but these would be far fewer and more likely to reflect real changes on the ground if the accuracy of the cartographic mapping process is high. Finally, although a pixel-based classification may help detect sub-object changes, pixel-based classifications have their own set of uncertainties and limitations (Liu and Xia, 2010). As previously described, image misregistration between pixels and the ‘salt-and-pepper’ effect become issues that must be addressed.

4.3.6 - Enabling long-term flexibility

Perhaps the most important consideration in long-term change detection studies is to ensure that the methodological framework remains flexible in order to remain meaningful over time and accommodate advances in remote sensing methodologies. Post-classification change detection between classified maps produced using different methodologies, such as survey-based thematic maps, EO-derived pixel-based and object-based classification maps, are often difficult to compare in a meaningful way (Comber et al., 2004; Fuller et al., 2003). However, retaining the original input imagery in a data storage system enables classifications to be re-run using different spatial frameworks as they are developed and updated (Comber and Wulder, 2019). Data storage and computer processing is no longer as large a constraint as it used to be; programmes such as the Land Change Monitoring, Assessment, and Projection (LCMAP) programme in the United States now produce annual land cover and change detection products using the entire Landsat image collection dating back to 1985 (Xian et al., 2022).

One option for Living England to consider is to change the spatial framework for each iteration of the habitat map but choose a single framework to remain consistent in the change detection analysis. Performing a new segmentation with the latest satellite imagery for each iteration of Living England ensures that each classification map contains image-objects relevant to that time-period. However, instead of performing post-classification change detection between two iterations of Living England with different spatial frameworks, a single set of image-objects is chosen from one iteration of Living England to act as a consistent spatial framework for change detection. This set of image-objects is then used to re-classify the input imagery for a different iteration of Living England. For example, Living England produces a habitat classification map in 2022 and 2024 and performs a segmentation for each of these two years. This generates two sets of image-objects, one that is derived from 2022 imagery and the other from 2024 imagery, which are classified using their respective year's imagery. Hypothetically, the 2024 image-objects are then chosen to be used in the change detection analysis. The 2022 imagery is then re-classified using the imposed 2024 image-objects. Post-classification change detection can then be performed between the 2024 classification map and the 2022 classification map (derived from the 2024 image-objects), with a consistent spatial framework.

This approach has the benefit of ensuring each iteration of Living England uses a set of image-objects relevant to the time-period, whilst enabling post-classification change detection to be performed with a consistent spatial framework. By retaining the input imagery, it enables long-term flexibility and opportunities for method advancement - any future updates to the Living England methodology could be applied to a previous year's imagery in the change detection process, reducing the possible errors that arise from methodological differences when comparing two classification maps. However, the change detection approach proposed here ultimately retains the same set of limitations associated with maintaining a spatial framework identified in Section 4.2.1. An appropriate method of detecting sub-object and object-boundary changes would need to be developed. It may be possible to detect sub-object changes and changes to object size and shape if two change detection maps were produced by this method. Returning to the previous example, comparing a change map produced using 2024 image-objects and one produced using 2022 image-objects may enable the detection of changes in object size and shape, but further research would be needed to determine if this is possible. Alternatively, a pixel-level classification summarised by objects may also help detect sub-object changes. Most importantly, a thorough and robust method for acquiring reference data for each time-period is still required to verify the accuracy of any changes detected. The difficulties in collecting reference data samples, the errors associated with post-classification change detection, and the possible approaches to improve classification accuracy will be discussed further in the following section.

4.4 - Accuracy and uncertainty in change detection

Evaluating the accuracy of any habitat change product is crucial to demonstrate its integrity and real-world applicability, yet land cover classification studies often fail to perform robust accuracy assessments on the final change maps which they produce (Olofsson et al., 2014). Accuracy assessments are more commonly performed on individual classification maps to assess their degree of 'correctness' compared to reference samples collected on the ground, but it is important to remember that all maps are the product of generalisations or models and will therefore contain some errors (Foody, 2002). To understand how these errors propagate when comparing two habitat classification maps in change detection, this section first provides an overview of the error matrix, the primary tool used in classification accuracy assessment. It discusses its limitations and describes the range of errors that can occur in the production of individual habitat classification maps. It then demonstrates how inaccuracies accumulate when performing post-classification change detection using the example of Land Cover Map 1990 and 2000 (Fuller et al., 2003). Finally, it suggests approaches adopted by studies to reduce these errors and improve the accuracy of change detection outputs.

4.4.1 - The error matrix and measures of accuracy

The primary tool used to assess classification accuracy is the error matrix, sometimes referred to as the confusion matrix, which compares the class labels from a classified map with the class labels obtained from a reference dataset (Olofsson et al., 2014). This enables the identification of two types of thematic errors: commission, and omission errors. Commission errors in classification describe when a reference data sample for a class is incorrectly identified as a different class by the classifier, whilst omission errors refer to reference data that are omitted from the correct class by the classifier. These errors are used to calculate two measures of accuracy known as user's and producer's accuracy (Figure 3). User's accuracy considers the end user's perspective of the habitat map and can be understood as the proportion of classified data in a class that actually belong to that class based on reference data. In Figure 3, of the 57 data objects classified as Forest, only 28 of these were actually Forest in the reference data, giving a user's accuracy of 49%. On the other hand, Producer's accuracy focuses on the producer's perspective and can be understood as the proportion of the reference data classified correctly in a given class (Story and Congalton, 1986). In Figure 3, of the 30 reference data samples for the Forest class, 28 of these were correctly classified, giving a producer's accuracy of 93%. Overall accuracy can then be calculated as the total number of reference data samples that have been correctly classified. Overall accuracy is typically used as the headline figure for the accuracy of maps, but as the examples in Figure 3 demonstrate, it is essential to recognise that classification accuracy is highly variable between different classes and between different measures of accuracy. This is an important message to convey to end users, as it can have important consequences for the practical application of a classification map.

Classified Data	Reference Data: F	W	U	Row Total
F	28	14	15	57
W	1	15	5	21
U	1	1	20	22
Column Total	30	30	40	100

Sum of the major diagonal = 63

Overall Accuracy = 63/100 = 63%

Producer's Accuracy

F = 28/30 = 93%
W = 15/30 = 50%
U = 20/40 = 50%

User's Accuracy

F = 28/57 = 49%
W = 15/21 = 71%
U = 20/22 = 91%

Figure 3: A hypothetical error matrix detailing the calculation of Overall, Users and Producers Accuracy for three classes: Forest, Urban and Water. The F:F, W:W and U:U boxes represent the number of correctly classified objects. Source: Story and Congalton (1986), reprinted with permission from the American Society for Photogrammetry & Remote Sensing, asprs.org.

4.4.2 - Limitations of the error matrix

Although the use of the error matrix is considered best practice for determining the accuracy of a given thematic map, it has led some studies to purely focus on reducing omission and commission errors by improving the classification algorithm (Foody, 2002). Although this is important to improve measures of accuracy, there are other limitations of error matrices that are often overlooked. For example, the accuracy and relevance of the reference dataset used in the error matrix is fundamental for determining the accuracy of the classification. Adequate amounts of reference data are required for each habitat class to validate the accuracy of the classifier in classifying that specific class. However, biases in data collection methods due to resources, the difficulty in identifying some habitats and even habitat accessibility can lead to some classes having much more reference data collected than others. Poorly chosen reference data that is incorrectly classified, mislocated or not spatially or temporally representative of that class can produce misleading accuracy figures (Olofsson et al., 2014). Living England is trying to ensure appropriate reference data is collected for each habitat by using bespoke ground truthing data records across the variety of habitats found in each biogeographic zone (Kilcoyne et al., 2022). Nonthematic errors also contribute to classification accuracy but are rarely captured in error matrices. These include errors in positional certainty and sensor properties, which can lead to the misregistration of classification and reference data, as well as the unequal spatial distribution of errors. However, error matrices are aspatial; they provide no information on the spatial distribution of errors, which are often concentrated in heterogenous landscapes and at the boundaries between classes (Comber, 2013).

Perhaps the biggest limitation of using the error matrix in classification is that it assumes that the unit of analysis (e.g. a pixel or an object) only belongs to one class, but this is often not the case (Foody, 2002). Hard classification approaches, which use discrete classes generated along a continuum, are likely to contain mixed pixels or objects due to the heterogenous nature of the landscape, particularly when using satellite input imagery with a resolution of 10 m. Since reference data is typically of a higher resolution (acquired through survey sampling for example), mixed pixels or objects classified as one class may therefore only actually be 'partially' correct or 'partially' incorrect, due to the mismatch in scale between the object and the reference sample. This contributes to variation in the magnitude and spatial distribution of classification errors, with errors being more likely to occur in heterogenous landscapes, such as the boundaries between habitats, than in homogenous ones. Reference sample points for Living England are collected in the field to specifically describe an image-object as defined by the segmentation. This addresses the differences in resolution between the reference data and the map product, but for heterogeneous objects this leaves the potential for differences in interpretation between the field surveyor and the classification algorithm. This approach also raises an issue when using the reference data in change detection, as object-boundary changes will lead to uncertainty as to whether the reference data is still relevant to the object or not.

4.4.3 - Error matrices in change detection

When applied to change detection, the error matrix becomes more complex. A change detection map produced via post-classification will identify transitions between classes and this map itself requires validation. To validate changed classes using the error matrix, reference data is required for each possible transition between classes. Gathering accurate, representative reference data for all possible transitions becomes extremely difficult due to the large number of possible changes that can occur ((Number of classes – $1)^2$) (Tewkesbury et al., 2015). If using field validation samples, a location needs to be repeatedly sampled each year to determine whether a change has occurred or not. Identifying which places may change class and what class they may change to is challenging, although a change taxonomy has recently been developed and so this may become more feasible in future (Lucas et al., 2022). Given the level of challenge this validation process is often done through manual visual interpretation of imagery, which is resource intensive and can still be prone to human error (Hazeu et al., 2011). This ideally requires consistent, well-trained photo-interpreters with knowledge of local ecological processes to accurately identify validation sites (Copernicus Land Monitoring Service, 2021). This process also needs to address the possible mismatch in resolution between the reference and the classified data, to ensure that it remains relevant for validating changes detected at a lower resolution (Foody, 2002). Validation using photo-interpretation has been widely used in national land cover mapping programmes to date (e.g. (Hazeu et al., 2011) and this remains a viable approach when resources constrain the ability to validate changes through field survey (Olofsson et al., 2014).

4.4.4 - Proliferation of errors in change detection

The reason independent reference data is required to validate changes is because errors present in individual maps are not independent and randomly distributed for a variety of reasons as outlined in Section 4.4.2. Hypothetically, if it is assumed that errors in a given classification map are independent and randomly distributed, then the accuracy of a change map could be determined by overlaying the two and multiplying the accuracies in each place (Fuller et al., 2003). However, in addition to identifying the correct change or stasis that has occurred, there would also be:

1. Areas mapped as ‘no-change’ which did in fact change on the ground
2. Areas mapped as changed, but the change detection process identified the wrong transition between classes. This could be either the result of an incorrect classification on one or both of the classification maps involved.
3. Areas mapped as changed when in fact no change has occurred.
4. Areas mapped as ‘no change’, but due to an incorrect classification in both of the maps involved.

Areas of change (c)		
	a_2	$1 - a_2$
a_1	**a_1a_2c change mapped correctly**	$a_1(1 - a_2)c(n - 2)/(n - 1)$ change mapped but incorrectly
		$a_1(1 - a_2)c/(n - 1)$ change hidden by chance error-match
$1 - a_1$	$a_2(1 - a_1)c(n - 2)/(n - 1)$ change mapped but incorrectly	$(1 - a_1)(1 - a_2)c(n - 2) / (n - 1)$ change mapped but ‘doubly incorrect’
	$a_2(1 - a_1)c/(n - 1)$ change hidden by chance error-match	*$(1 - a_1)(1 - a_2)c/(n - 1)$ change hidden by chance error-match*

Areas with no change (1 – c)		
	a_2	$1 - a_2$
a_1	**$a_1a_2(1 - c)$ stasis mapped correctly**	$a_1(1 - a_2)(1 - c)$ stasis mapped as change
$1 - a_1$	$a_2(1 - a_1)(1 - c)$ stasis mapped as change	$(1 - a_1)(1 - a_2)(1 - c)(n - 2)/(n - 1)$ stasis mapped as change
		$(1 - a_1)(1 - a_2)(1 - c)/(n - 1)$ stasis mapped as such but through double errors

The notation is as follows: a_1 is the proportional accuracy of map-1; a_2 is the proportional accuracy of map-2; c is the proportional change between mapping-dates; n is the number of classes mapped. Note that accuracies and errors are multiplicative; also that, where there is change, there is a chance $(1/(n - 1))$ that an error may hide the change; where there is stasis, all single-date errors appear as change but double-date errors have a chance of recording stasis, albeit of the wrong class. Where the mapping shows the true situation (whether change or stasis) the cell text is shown in bold; true changes hidden by errors are given in italic text; where the mapping records a difference (whether it is change or error), the cells are shaded.

Figure 4: Equations used to compare pairs of maps to assess differences due to change and those due to errors; also to identify changes hidden by errors. Source: Fuller et al. (2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier.

The overall changes identified in a change map can thus be understood as the sum of real changes plus real changes recorded incorrectly and static situations where a change was recorded due to an error in either or both of the input classification maps (Figure 4). This is best demonstrated in a paper by Fuller et al. (2003), who hypothetically analysed changes between Land Cover Map 1990 and Land Cover Map 2000 for the UK. The overall accuracy for LCM 1990 was 80%, whilst the accuracy for LCM 2000 was 85% - considered as high accuracies by the remote sensing community. The Countryside Survey data used to calibrate the LCMs identified an overall change of 17% in the primary codes used to record major land cover types. When using the mathematical equations above, Fuller et al. (2003) noted that there would only be a 57% agreement between the two maps, meaning that 43% of the map area would record differences (Figure 5). 17% would reflect the real changes identified in the Countryside Survey data, but that leaves a further 26% reflecting changes arising through errors. Although they estimated 98% of actual changes would be captured, this is largely due to a significant overestimation of the changes occurring, with many false positive changes being mapped.

Proportional accuracy of LCMGB:	a_1	0.80
Proportional accuracy of LCM2000:	a_2	0.85
Indicative proportion of change:	c	0.17
Number of classes:	n	16

Areas of change (c)			
	a_2	$1\text{-}a_2$	**Totals**
a_1	**0.12**	0.02	0.13
		0.00	0.00
$1\text{-}a_1$	0.03	0.00	0.03
	0.00	*0.00*	0.00
Totals	0.14	0.03	0.17

Areas with no change (1-c)			
	a_2	**$1\text{-}a_2$**	**Totals**
a_1	**0.56**	0.10	0.66
$1\text{-}a_1$	0.14	0.02	0.16
		0.00	0.00
Totals	0.71	0.12	0.83

	proportional coverage	percent of map area
Maps show different classes (sum of shaded cells)	0.43	43%
Maps show the same classes (sum of unshaded cells)	0.57	57%
Real change as a proportion of mapped difference	0.40	40%
Proportion of change which is correctly shown as such	0.98	98%

Where the mapping shows the true situation (whether change or stasis) the cell text is shown in bold; true changes hidden by errors are given in italic text; where the mapping records a difference (whether it is change or error), the cells are shaded.

Figure 5: Assessing likely differences due to change and those due to errors between Land Cover Map 1990 and 2000. Source: Fuller et al. (2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier.

Fuller et al. (2003) conclude that based on these equations the level of accuracy needed in each map to measure this change with 75% reliability (75% of observed differences between maps are actual changes) would need to exceed 95% (Figure 6). Importantly, the higher the number of habitat classes in each classification map and the smaller the proportion of the area of expected change, the higher the accuracy required for each classification map. Living England contains 17 detailed classes and aims to perform change detection between each iteration produced every two years. The proportion of real changes likely to occur over such a time-period is therefore very small, probably much less than the 17% identified between LCM 1990 and 2000. This means that in order to reliably detect (>75% reliability) the small number of real changes expected to occur, the individual Living England map accuracies would need to be almost 100%. Otherwise, any changes detected could be due to errors or chance, rather than reflecting real changes occurring.

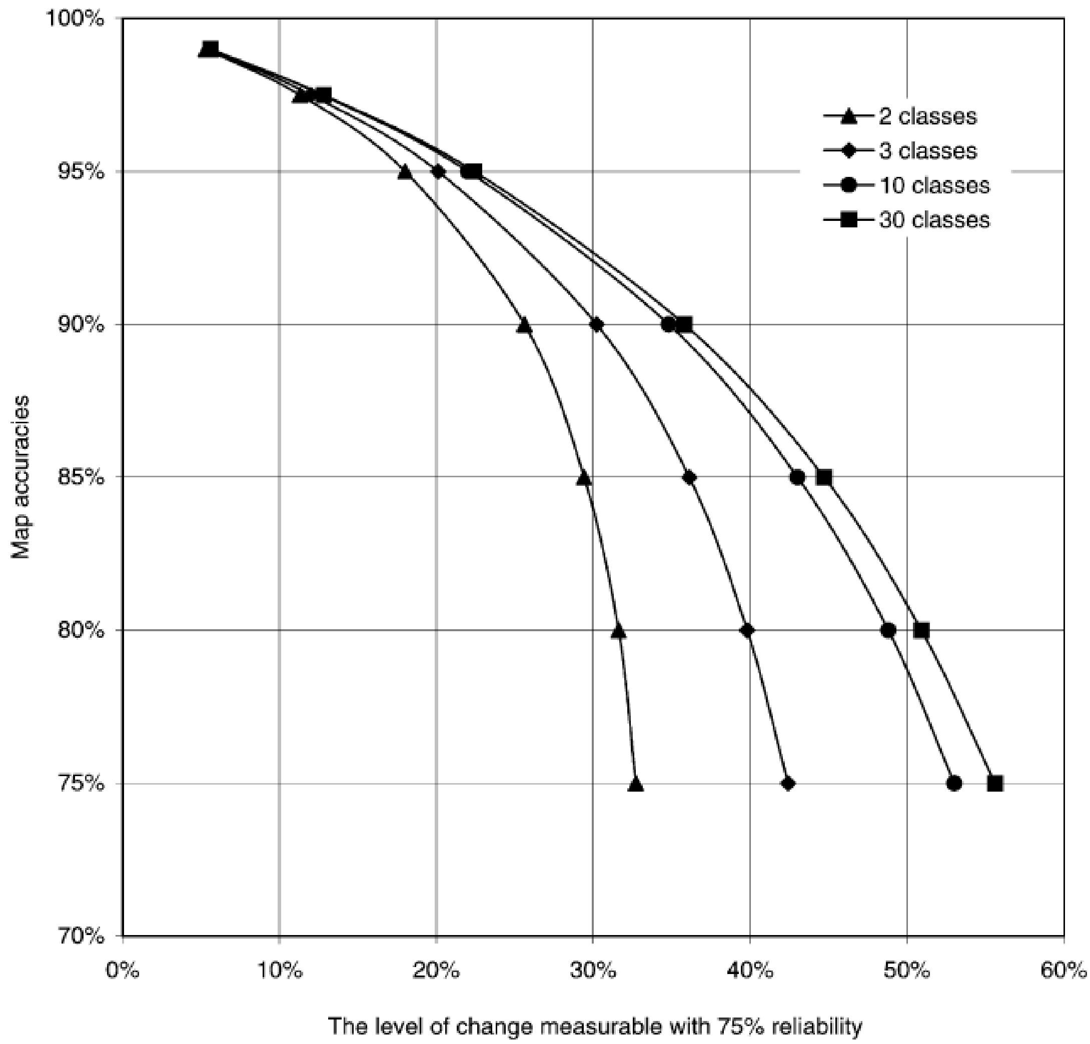

Figure 6: The proportional area of expected change which can be measured with 75% reliability when using pairs of maps with varying class numbers and map accuracies. As the proportional area of expected change declines, the higher the accuracy of the maps required in order to reliably detect the change. Source: Fuller et al .(2003), this article was published in Int. J. Appl. Earth Obs. Geoinformation 4, 243-253, Copyright Elsevier.

It is important to remember that the example above assumes that there is independence in the spatial distribution and magnitude of errors, but as discussed, errors in classification exercises often vary in magnitude and can be spatially autocorrelated (Comber, 2013). It also ignores any errors arising from the misregistration of spatial frameworks/images, differences in the methodologies applied (LCM 1990 was produced using a different classification approach than LCM 2000) and errors arising from thematic differences in the maps themselves, such as the definitions of each class (Comber et al., 2004). Therefore, these additional sources of error and uncertainty are likely to reduce the true accuracy of classified maps, and any derived change products, even further.

4.5 - Approaches to support accuracy assessment in change detection

Although accuracy assessment in post-classification change detection is complex, there are numerous methods that can be adopted to make this task more feasible. These include reducing the number of transition classes being examined and adopting a thorough and suitable sampling strategy. However, the most important consideration that all these approaches share is the need to clearly define the objectives and requirements of the change detection product. This section first discusses this consideration, before describing some of the possible approaches used to assist and communicate accuracy assessment in post-classification change detection.

4.5.1 - Clear objectives for change detection

The importance of having clear objectives and end-user requirements for determining a suitable change detection methodology has been emphasised widely throughout the literature (Comber and Wulder, 2019; Hall and Hay, 2003; Lightfoot et al., 2020; Smith and Morton, 2010; Tewkesbury et al., 2015). The scale, thematic and end-user requirements of the change detection product will have important implications for the types of change able to be examined, the accuracy measures to be prioritised and ultimately the selection of the most appropriate change detection technique. For example, a national policy advisor may need to report on national-scale habitat changes and therefore prioritises overall accuracy in the change map. Conversely, a local conservation group working on a Nature Recovery Network may need to accurately monitor the dynamics of a rare habitat; they would prioritise the accuracy of a particular set of relevant transitions and want to ensure that no changes are missed (low omission errors). Whilst this is just a hypothetical example, it highlights the importance of clearly understanding the end-users' requirements to generate a useful and applicable change detection product.

From a spatial scale perspective, it is important that the segmentation scale and image resolution adequately reflect the features of interest and the scale of change expected to occur (Hall and Hay, 2003); regions of change in segments must represent a significant proportion of the object in order to be detected. If the objective is to identify large changes in the landscape, then classifying image-objects is likely more appropriate than pixels. However, if landowners need to demonstrate that they have implemented a habitat creation option under the new ELM schemes, then identifying sub-object changes is important and so a pixel-based approach may be more appropriate, otherwise higher resolution mapping or field visits would be required. Consideration should also be given to the temporal scale of change detection. Ensuring the temporal scale of change detection is aligned to the expected temporal scale of ecological changes is important to ensure accurate monitoring (Comber and Wulder, 2019). (Lightfoot et al., 2020) examined annual intertidal habitat dynamics using OBIA and found that large areas of change were being detected on an annual basis (approximately 13-40% depending on the sensitivity of the change detection method). However, a lot of these changes reflected the short-term, cyclical nature of some ecological processes, rather than long-term habitat change occurring over the study period. The temporal scale of observations needs to match the type of change being monitored; Living England may therefore want to consider whether to assess change at a range of intervals to identify longer-term patterns of change as opposed to just short-term changes between iterations which may be less meaningful or result in unnecessary management intervention.

Some types of change occur over very short time periods due to management, weather events or seasonal variation. These changes may impact on classification accuracy, but also require a different approach to assessing change, which is why a related JNCC project is delivering monthly time series satellite data over several years related directly to the Living England habitat map. Such data will enable users to review change and assess impacts of these changes within habitats rather than assessing the level of complete change where a location had changed from one class to another, which is what is being described in this work. For a more detailed description of this work see Lightfoot et al. (2021).

4.5.2 - Designing an appropriate sampling strategy

Once clear objectives for the change detection product have been determined, appropriate approaches for accuracy assessment can be chosen (Olofsson et al., 2014). If using the error matrix described above, collecting reference data for transitions is difficult and time consuming, but there are numerous ways that this can be simplified. First, specific transitions can simply be ignored if they are not considered as important for the objectives of the change detection product, for example between urban and water classes. Equally it is probable that some transition classes can be removed because they are unlikely to occur in certain geographic regions – the transition between coastal saltmarsh and upland bog for example (Comber et al., 2004). Transitions could also be removed if they contain a class which has low classification accuracies in either of the classification maps, due to the issues of error multiplication described in Section 4.4.4. The targeting of transitions could be informed by the Global Change Taxonomy for Land Cover devised by Lucas et al. (2022) and the version that has been produced specifically to describe change in UK habitats (Lucas et al., in prep). These provide details of the likely changes that can impact on specific habitats or land cover types and relates them to pressures and drivers that cause those changes. Living England has already taken steps towards this, by running a separate classification in Phase III using only classes which have greater than 45 training points – achieving significantly higher accuracies (Kilcoyne et al., 2021). Limiting transitions to classes with higher accuracy and prioritising transitions which are considered important for monitoring purposes can all reduce the amount of reference data that needs to be collected and enable more robust change detection.

Another approach to reduce the number of transition classes is to aggregate more detailed classes together before performing change detection. This can often improve the classification accuracy of the input maps, by combining difficult to distinguish classes (e.g.

acid grassland from calcareous grassland), as well as reducing the number of possible transitions, making validation of the change detection output easier. This approach has been adopted in the Netherlands for their national land cover mapping programme. Thirty nine classes were aggregated into eight for the change detection process, based on those considered to be the most socio-economically relevant (Hazeu et al., 2011). However, if there are a range of end-users then several change maps could be produced representing varying 'sensitivity' levels with differing thresholds for change (Lightfoot et al., 2020). Lightfoot et al. (2020) used three sensitivity levels in their study of intertidal habitats: High included all possible transitions between classes (except shadow and strand-line), medium represented all transitions except between algal classes and low represented only transitions between algal classes and non-vegetated surfaces. All maps achieved overall accuracies above 75%, with the low sensitivity map achieving the highest overall accuracy due to the reduced number of transitions being identified between spectrally similar classes. Conversely, the high sensitivity map tended to over-classify change, leading to lower user's accuracy for transitions (high commission errors). Although overall accuracy for each of the three maps was similar (between 70% and 80%), the proportion of the total area identified as 'change' varied between 15 and 40% (Figure 7, Lightfoot et al., 2020). This highlights the importance of presenting the full error matrix, with guidance, for end-users to interpret, as well as prioritising the type of accuracy to be achieved. Focusing on identifying all possible changes and prioritising overall accuracy figures is not always recommended in change detection and the sampling strategy adopted should reflect this (Olofsson et al., 2014).

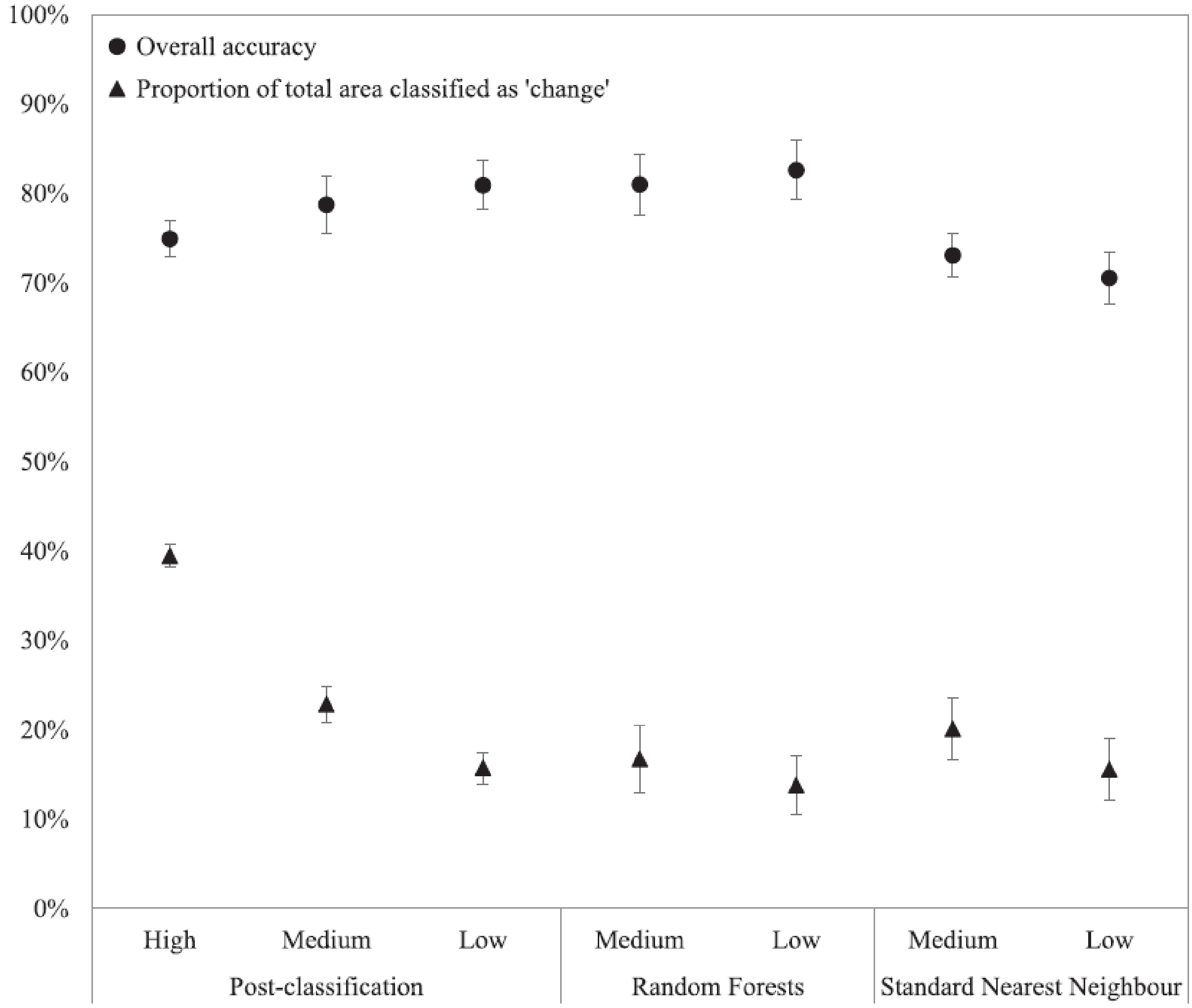

Figure 7: Mean overall accuracy and mean proportion of total area classified as 'change' in high, medium, and low sensitivity change detection maps created by object-based analysis of multi-temporal aerial and LiDAR imagery (*n* = 13). Error bars represent standard deviation. Source: Lightfoot et al. (2020)

The sampling strategy used to gather reference data for training and validation is a crucial step in accuracy assessment and should be tailored to the objectives of the study (Olofsson et al., 2014). Whilst there are some design principles that remain consistent – the sampling strategy should include randomisation; the reference data should be of higher quality than the classified data and the validation data must be independent to the classifier's training data – some elements of the sampling strategy can be designed to prioritise specific measures of accuracy. Depending on the relative importance of omission v commission errors of specific classes to the end user, a sampling strategy may prioritise collecting more samples for difficult to identify transitions/classes to improve their accuracy. Olofsson et al. (2014) recommend increasing the sample size for rare classes to achieve acceptable accuracies, before allocating the remaining sample size proportionally to the remaining classes based on their area, although this is more challenging to collect when applied to transition classes in a change detection map. They also suggest performing sampling in sub-regions of the area to be classified, to reduce the influence of sub-regional errors on the wider classification. However, it is also important to recognise that different classifiers respond to training data differently. The Random Forest classifier is acknowledged to be sensitive to the proportions of classes used in training samples, so using an area-proportional sampling strategy may produce more accurate results than one which uses an equal-allocation design (Lightfoot et al., 2020; Olofsson et al., 2014). Further guidance on how to design an appropriate sampling strategy to meet specific objectives is given in Olofsson et al. (2014).

4.5.3 - Communicating and visualising uncertainty

Irrespective of the accuracy of change detection, it is important to communicate and visualise this uncertainty to the end-user for the outputs to be used appropriately. As described previously, errors in classification and change detection can vary in magnitude between classes and not be randomly distributed across the landscape (Comber, 2013; Foody, 2002). This is particularly true in change detection, where specific transitions can be hard to distinguish, and their validation sites may be localised due to context-specific drivers of change.

Methods to map spatial uncertainty and accuracy in classification maps are well established (Robinson, 2007; Zhang and Goodchild, 2002). For example, JNCC have previously visualised spatial uncertainty by mapping the Random Forest probabilities of the most likely class to denote the certainty with which the classifier has classified a specific object (Colson et al., in press). Furthermore, they also mapped the % difference between the top two classes identified by the Random Forest to highlight areas where the classifier is more uncertain about the most likely habitat class (Colson et al., in press). Although difficult to meaningfully translate to a post-classification change detection map, since transitions are not directly classified, this approach could be employed to identify areas on the input classification maps where the uncertainty between classes is too high to perform meaningful change detection in that area.

In addition to mapping classification uncertainty, (Comber, 2013) has developed geographically weighted logistic regression models to estimate local scores of overall, user's and producer's accuracy measures for different classes (Figure 8). By developing a logistic model to represent each of these accuracy measures for a given class, these can be applied in different localities using a kernel or moving window to make local calculations using the data points that are under the kernel, but whose contribution to the calculation is weighted by their distance (Comber, 2013). The kernel is based on a grid, giving location specific estimates of the accuracy of each class in that specific location. Although not yet applied to a change detection map, geographically weighted logistic regression could feasibly be extended to a change detection product, given its use of overall, user's and producer's accuracy, which would be determined if using the error matrix to validate the change detection map. It should be noted that a spatially distributed sample of reference data points is still required for this, as well as consideration of the number of data points to be included in the local model (Comber, 2013). However, this approach would be particularly helpful for local end-users, who can determine whether the change detection accuracy is high enough in their local area for a specific transition that they are interested.

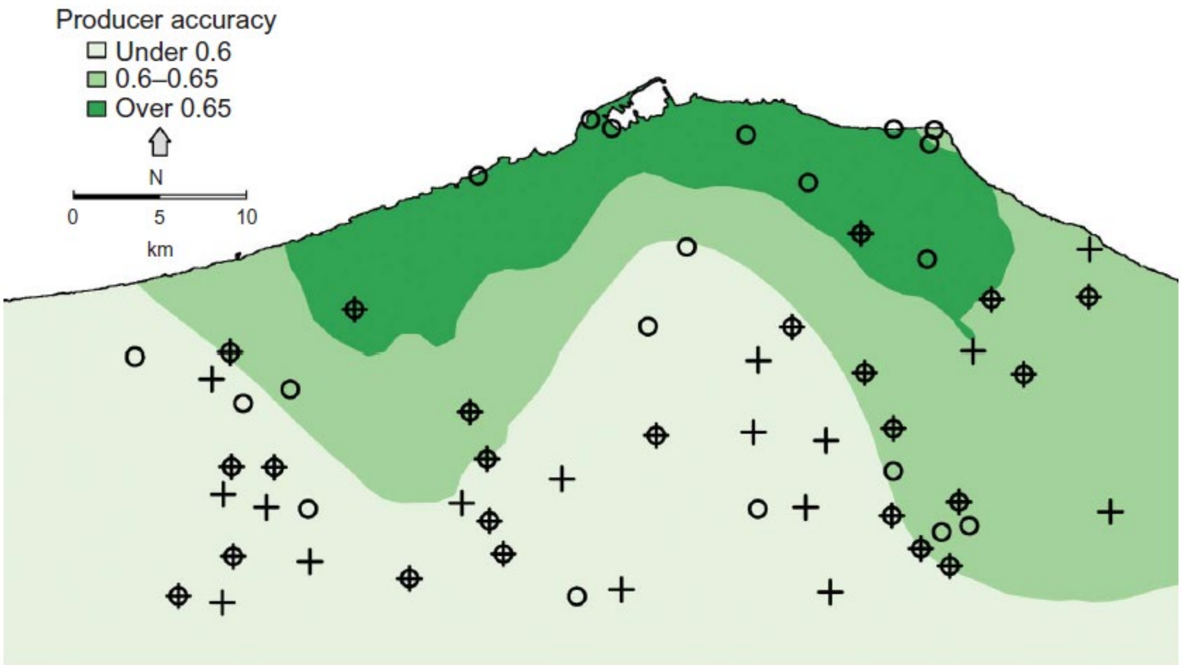

Figure 8: An example of the spatial variation in producer accuracy for the 'Grazing Land' class. Crosses indicate locations where Grazing Land was recorded in the reference data, whilst circles indicate locations where Grazing Land was recorded in the classification. Source: Comber (2013)

If the change detection estimates are used for national-scale monitoring of habitat areas, then communicating uncertainty in the extent of habitat changes would be required. This is more challenging to assess, but some classification studies have done this by assessing the proportion of a habitat class area that has been classified over a certain probability. This can be presented graphically and divided into equal intervals (alpha-cuts) to show the geographical area by probability for each given class (Fisher, 2010). For example, the top 10% alpha-cut would represent the proportion of the habitat class area classified with a probability exceeding 90%. The second alpha-cut would show the proportion of the habitat class area classified with a probability between 80-90%. Alternatively, JNCC categorised polygons by their area and then averaged the probability of the most likely class for each polygon in each size category, demonstrating that the smaller the polygon, the lower the average probability for a given class (Colson et al., in press). Whilst these techniques can help to identify classes and locations where the classifier is uncertain, potentially removing spurious transitions, they are generated directly from the Random Forest output, and it is not clear how these could be adopted for a change detection output. One option may be to summarise the geographically weighted surfaces of accuracy for each transition class by polygon size or geographic region, but this would require further research.

Alternatively, there may be fuzzy classification approaches which could generate a range of possible area estimates for a given class or transition, but these would need to be researched in further detail and would likely require a significant change to the Living England methodology (Fisher, 2010). If fuzzy classification algorithms are used, it is important to note that alternative approaches to accuracy assessment would be needed, as the error matrix implicitly assumes that objects belong to only one class and is therefore not applicable (Feizizadeh et al., 2017). This is because fuzzy approaches are describing vagueness in the objects and their definition rather than randomness in error (Fisher, 1999; Zhang and Goodchild, 2002). The Living England process does not use a fuzzy Random Forest algorithm, meaning that not all the attributes (i.e. input datasets) and classes are considered for all decision points as the algorithm creates a decision tree. For this reason, the probability values produced cannot be considered as fuzzy membership values, but they are a measure of certainty that the class assignment is true, so it is possible to utilise them for analysing uncertainty in the classification as described above.

4.5.4 - Manual identification of changes

Finally, it should be briefly noted that due to the difficulties associated with error propagation in post-classification change detection, and the importance of achieving highly accurate results; most national-scale programmes resort to the manual identification of changes using photo-interpretation and field surveys (Copernicus Land Monitoring Service, 2021; Hazeu et al., 2011). This is both resource intensive and time consuming but produces more accurate and reliable results. Hazeu et al. (2011) identified binary change/no change between eight land cover classes across the Netherlands in 50 days through photo-interpretation. Similarly, the CORINE programme produces a change layer across Europe using specialised software and trained photo-interpreters to directly classify and code the changes being observed using a manual of changes. The two images used in the production of the classified products are compared in a dual-window environment, with the old polygons used to delineate changes and prevent the production of sliver polygons. Polygons less than 100 m in width and 5 ha in size are removed from the process and the change layer is subsequently validated through independent re-interpretation of the imagery, assisted with higher resolution Google Earth images (Copernicus Land Monitoring Service, 2021). The method is resource intensive but achieves accuracies for each change type of around 85%, significantly higher than could ever be achieved through post-classification alone (Foody, 2002). Given that validation of changes often requires image interpretation already, this could be considered as an option depending on the resources available and the end users of the Living England products.

5 - Recommendations

This review has identified numerous issues associated with OBCD and possible approaches to address some of these issues. Based on this, we suggest the following set of recommendations to be considered by Living England for the development of their change detection methodology.

5.1 - General recommendation

1. **Clearly define the objectives of the change detection product based on the scale, thematic and end-user requirements of the outputs**. This is a fundamental first step in change detection and is crucial for the selection of an appropriate unit of analysis, the right change detection methodology and the approach to accuracy assessment. This should include, but not be limited to, the following questions:
    a. **Who are the end users?** (E.g. Reserve managers, ELM applicants, policy-makers)
    b. **What type and scale of habitat changes are they most interested in?** (E.g. national v local, object-level v sub-object level)
    c. **Can these changes be feasibly detected via Earth observation?** (Examine the Habitat Change Detection Framework - (Lucas et al., in prep))
    d. **Do these changes align with the spatial and temporal scale of the Living England approach**? It is important that the segmentation scale and image resolution adequately reflect the features of interest and the scale of change expected to occur (Hall and Hay, 2003); regions of change in segments must represent a significant proportion of the object in order to be detected.
        i. **If not, can the Living England approach be easily adapted to meet this purpose?** (i.e. classifying by pixels and summarising by objects to detect sub-object changes or examining longer-term change between successive LE iterations)

5.2 - Post-classification change detection

1. Maintain the spatial framework from one of the input maps and use this to analyse imagery from both time points to avoid the production of sliver polygons. Sliver polygons are difficult to practically manage and can arise from methodological inconsistencies or reflect real changes. Simply removing these polygons based on an area threshold may remove valuable information on small-scale habitat dynamics and so this is not recommended. Although image-objects lose their relevance through time, maintaining the spatial framework is more practical for change detection.
2. **Retain all the unclassified input imagery to enable flexibility when performing post-classification change detection.** A new segmentation can be performed for each Living England iteration; creating temporally relevant

image-objects and incorporating methodological advances that may occur. By retaining the unclassified input imagery, the spatial framework from any one iteration can then be used to re-classify imagery from any other previous iteration, enabling change detection to be performed with a consistent spatial framework across any time interval between iterations. This also has the benefit of avoiding the production of sliver polygons.

3. **Reduce the number of transition classes being detected**. This reduces the amount of reference data collection that needs to be performed and will improve the accuracy of the change detection output. If applicable, use sensitivity maps to present a range of change-detection outputs with differing thresholds for detecting change, such as the number and type of transitions captured. Consider either aggregating similar classes together or removing transition classes that:

    a. Are not relevant to the objectives of the output (ecologically, geographically, economically)
    b. Involve classes that have low levels of classification accuracy in either input map
    c. Occur in areas where there are high levels of uncertainty in the classification algorithm. Assess this using either the % probability of the most likely class or the degree of confusion between classes (Colson et al., in press)

4. **Design a robust sampling strategy to independently validate transitions identified in the change detection product.** Ensure that this reflects the objectives and priorities of the change detection product; follow the guidance of Olofsson et al. (2014) where possible. Collect independent, randomly distributed reference data for each sub-region if resources are available. Determine whether to use an equal or area-proportional allocation of validation data to each transition class based on the priorities for accuracy assessment.
5. **Present more than just overall accuracy figures for the change detection product.** Include details of user's and producer's accuracies for each transition and ideally present maps of the spatial variation in these accuracies. Ensure that full details of the sampling strategy and the original error matrix are made available and that there is accompanying guidance for end-users to follow.

5.3 - Consider alternatives

1. Consider the use of alternative change detection and accuracy assessment approaches if post-classification change detection is not well suited to the user objectives. It is important to acknowledge that post-classification object-based change detection may not be the most suitable approach for detecting habitat changes, depending on the user requirements. For example, if accurately detecting all possible changes is a priority, then consider manually identifying changes via photo-interpretation (e.g. CORINE and (Hazeu et al., 2011). If sub-object changes are a priority, then pixel-based approaches, such as classifying

by pixel and summarizing by a digital cartographic framework, may be more suitable (Smith and Morton, 2010). Studies in North America frequently use a continuous change detection algorithm based on best-pixel composites, although the landscape is quite different to the UK (Hermosilla et al., 2016; Xian et al., 2022). Finally, objects with high degrees of confusion between classes may benefit from fuzzy approaches to classification and change detection (Fisher, 2010).

6 - References

Blaschke, T., 2005. Towards a framework for change detection based on image objects. Gött. Geogr. Abh. 113.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. ISPRS J. Photogramm. Remote Sens. 87, 180–191. https://doi.org/10.1016/j.isprsjprs.2013.09.014

Bontemps, S., Bogaert, P., Titeux, N., Defourny, P., 2008. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. Remote Sens. Environ. 112, 3181–3191. https://doi.org/10.1016/j.rse.2008.03.013

Colson, D., Hecker, L., Lightfoot, P., Robinson, P., in press. Northern Ireland Habitat Mapping Project: Mapping Fermanagh (No. JNCC Report 683). JNCC, Peterborough.

Comber, A., Fisher, P., Wadsworth, R., 2004. Assessment of a Semantic Statistical Approach to Detecting Land Cover Change Using Inconsistent Data Sets. Photogramm. Eng. Remote Sens. 70, 931–938. https://doi.org/10.14358/PERS.70.8.931

Comber, A., Wulder, M., 2019. Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. Trans. GIS 23, 879–891. https://doi.org/10.1111/tgis.12559

Comber, A.J., 2013. Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. Remote Sens. Lett. 4, 373–380. https://doi.org/10.1080/2150704X.2012.736694

Copernicus Land Monitoring Service, 2021. CORINE Land Cover Product User Manual Version 1 (No. Version 1). European Environment Agency (EA), European Union (EU).

DEFRA, 2022. Biodiversity Terrestrial and Freshwater Targets (Detailed Evidence Report). Department for Environment, Food and Rural Affairs.

DEFRA, 2021. Outcome Indicator Framework for the 25 Year Environment Plan: 2021 Update. Department for Environment, Food and Rural Affairs.

Desclée, B., Bogaert, P., Defourny, P., 2006. Forest change detection by statistical object-based method. Remote Sens. Environ. 102, 1–11. https://doi.org/10.1016/j.rse.2006.01.013

Feizizadeh, B., Blaschke, T., Tiede, D., Moghaddam, M.H.R., 2017. Evaluating fuzzy operators of an object-based image analysis for detecting landslides and their changes. Geomorphology 293, 240–254. https://doi.org/10.1016/j.geomorph.2017.06.002

Fisher, P.F., 2010. Remote sensing of land cover classes as type 2 fuzzy sets. Remote Sens. Environ. 114, 309–321. https://doi.org/10.1016/j.rse.2009.09.004

Fisher, P.F., 1999. Models of uncertainty in spatial data, in: Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (Eds.), Geographical Information Systems: Volume 1 Principles and Technical Issues. Wiley and Sons, New York, pp. 191–205.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Fuller, R.M., Smith, G.M., Devereux, B.J., 2003. The characterisation and measurement of land cover change through remote sensing: problems in operational applications? Int. J. Appl. Earth Obs. Geoinformation 4, 243–253. https://doi.org/10.1016/S0303-2434(03)00004-7

Hall, O., Hay, G.J., 2003. A Multiscale Object-Specific Approach to Digital Change Detection. Int. J. Appl. Earth Obs. Geoinformation 4, 311–327. https://doi.org/10.1016/S0303-2434(03)00010-2

Hayhow, D., Eaton, M., Stanbury, A., Burns, F., Kirby, W., Bailey, N., Beckmann, B., Bedford, J., Boersch-Supan, P., Coomber, F., Dennis, E., Dolman, S., Dunn, E., Hall, J., Harrower, C., Hatfield, J., Hawley, J., Haysom, K., Hughes, J., Johns, D., Mathews, F., McQuatters-Gollop, A., Noble, D., Outhwaite, C., Pearce-Higgins, J., Pescott, O., Powney, G., Symes, N., 2019. The State of Nature 2019. The State of Nature Partnership.

Hazeu, G.W., Bregt, A.K., de Wit, A.J.W., Clevers, J.G.P.W., 2011. A Dutch multi-date land use database: Identification of real and methodological changes. Int. J. Appl. Earth Obs. Geoinformation 13, 682–689. https://doi.org/10.1016/j.jag.2011.04.004

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int. J. Digit. Earth 9, 1035–1054. https://doi.org/10.1080/17538947.2016.1187673

Kilcoyne, A., Clement, M., Moore, C., Picton Phillipps, G., Keane, R., Woodget, A., Potter, S., Stefaniak, A., Trippier, B., 2022. Living England: Satellite-based Habitat Classification. Technical User Guide (No. NERR108). Natural England.

Kilcoyne, A., Cox, P., Picton Phillipps, G., Tomline, N., Keane, R., 2020. Living England: Satellite-based Habitat Classification. Phase 2 (No. Phase 2). Natural England.

Kilcoyne, A., Picton Phillipps, G., Keane, R., Clement, M., Moore, C., Woodget, A., 2021. Living England: Satellite-based Habitat Classification. Phase 3 (No. Phase 3). Natural England.

Lawton, J., Brotherton, P., Brown, V., Elphick, C., Fitter, A., Forshaw, J., Haddow, R., Hilborne, S., Leafe, R., Mace, G., Southgate, M., Sutherland, W., Tew, T., Varley, J., Wynne, G., 2010. Making Space for Nature: a review of England's widlife sites and ecological network. Report to Defra.

Lightfoot, P., French, G., Hassall, I., Hecker, L., Jones, A., Trippier, B., Robinson, P., 2021. Using Sentinel data to monitor change in habitats and historic landscape features. Technical Report for the Caroline Herschel Framework Partnership Agreement for Copernicus User Uptake (Work Package Six) (No. JNCC Report 687). JNCC, Peterborough.

Lightfoot, P., Scott, C., Fitzsimmons, C., 2020. Using object-based image analysis with multi-temporal aerial imagery and LiDAR to detect change in temperate intertidal habitats. Aquat. Conserv. Mar. Freshw. Ecosyst. 30, 514–531. https://doi.org/10.1002/aqc.3277

Liu, D., Xia, F., 2010. Assessing object-based classification: advantages and limitations. Remote Sens. Lett. 1, 187–194. https://doi.org/10.1080/01431161003743173

Liu, T., Yang, L., Lunga, D., 2021. Change detection using deep learning approach with object-based image analysis. Remote Sens. Environ. 256, 112308. https://doi.org/10.1016/j.rse.2021.112308

Lucas, R., Hurford, C., Peck, K., Planque, C., Punalekar, S., Robinson, P., in prep. Assessing the Use of Earth Observation Data for Habitat Change: A UK Demonstration.

Lucas, R.M., German, S., Metternicht, G., Schmidt, R.K., Owers, C.J., Prober, S.M., Richards, A.E., Tetreault-Campbell, S., Williams, K.J., Mueller, N., Tissott, B., Chua, S.M.T., Cowood, A., Hills, T., Gunawardana, D., McIntyre, A., Chognard, S., Hurford, C., Planque, C., Punalekar, S., Clewley, D., Sonnenschein, R., Murray, N.J., Manakos, I., Blonda, P., Owers, K., Roxburgh, S., Kay, H., Bunting, P., Horton, C., 2022. A globally relevant change taxonomy and evidence-based change framework for land monitoring. Glob. Change Biol. 28, 6293–6317. https://doi.org/10.1111/gcb.16346

Morton, R.D., Rowland, C.S., 2015. Developing and Evaluating an Earth Observation-enabled ecological land cover time series system (No. JNCC Report 563). JNCC, Peterborough.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

Robinson, P., 2007. Evaluating Uncertainty in Classification within the Land Cover Map 2000 (PhD). University of Leicester, Leicester.

Shepherd, J.D., Bunting, P., Dymond, J.R., 2019. Operational Large-Scale Segmentation of Imagery Based on Iterative Elimination. Remote Sens. 11, 658. https://doi.org/10.3390/rs11060658

Smith, G.M., Morton, R.D., 2010. Real World Objects in GEOBIA through the Exploitation of Existing Digital Cartography and Image Segmentation.

Photogramm. Eng. Remote Sens. 76, 163–171. https://doi.org/10.14358/PERS.76.2.163

Song, A., Kim, Y., Han, Y., 2020. Uncertainty analysis for object-based change detection in very high-resolution satellite images using deep learning network. Remote Sens. 12. https://doi.org/10.3390/RS12152345

Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. Photogramm. Eng. Remote Sens. 52, 397–399.

Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., Fisher, P.F., 2015. A critical synthesis of remotely sensed optical image change detection techniques. Remote Sens. Environ. 160, 1–14. https://doi.org/10.1016/j.rse.2015.01.006

Xian, G.Z., Smith, K., Wellington, D., Horton, J., Zhou, Q., Li, C., Auch, R., Brown, J.F., Zhu, Z., Reker, R.R., 2022. Implementation of the CCDC algorithm to produce the LCMAP Collection 1.0 annual land surface change product. Earth Syst. Sci. Data 14, 143–162. https://doi.org/10.5194/essd-14-143-2022

Zhang, J., Goodchild, M.F., 2002. Uncertainty in Geographical Information. CRC Press, London. https://doi.org/10.1201/b12624

Zhang, X., Xiao, P., Feng, X., Yuan, M., 2017. Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area. Remote Sens. Environ. 201, 243–255. https://doi.org/10.1016/j.rse.2017.09.022

Annex 1: Literature Review process

As described in the methodology, a systematic literature review process was not adopted due to the time constraints associated with the review. Instead, the literature review focused on the use of a known, relevant, and recent review articles on the subject matter identified by NE; literature identified and recommended from the expert interviews; and a time-limited literature search of the Scopus database for further literature. Details of the methodology used in the time-limited literature search are described below.

Annex 1.1 - Search String Development

Due to the time constraints identified above, a thorough scoping exercise for relevant search terms/strings was deemed impractical. Instead, the authors made use of the DEFRA Library Services Team, who can develop relevant search strings upon request. The Library team were requested to perform a targeted literature search on the topic, using a set of identified key words, the existing literature list provided by NE and a description of the aims of the literature review. As part of this request, literature was narrowed down to peer-reviewed articles published in the last 20 years and those which were published in English.

The DEFRA Library Team struggled to narrow down the search to under 750 hits with their initial search strings. The primary author took the initial string from DEFRA and made several alterations to narrow down the number of hits being returned from literature searches. Each string was recorded and assessed for its ability to return relevant hits using the initial list of literature supplied by NE and further list of articles identified from two recent review articles. Two search strings were then taken forward to compile the literature for screening, based on having the highest number of combined matches between them. One was a broad search string of titles (153 hits) and the second based on Titles, Abstracts and Keywords (207 hits). In total, 291 different articles were captured. This was done to try and focus specifically on methodological papers (titles string), but also trying to capture any papers that refer to a segmentation approach, change detection accuracy assessment or uncertainty analysis etc, but not explicitly reference them in the title. Erratums and corrections to articles were removed from the reference library.

Title, Abstract and Keywords:

TITLE-ABS-KEY (("IMAGE SEGMENTATION" OR SEGMENTATION) AND (POST-CLASSIFICATION OR "CHANGE DETECTION") AND OBJECT-BASED) AND (LIMIT-TO (EXACTKEYWORD , "IMAGE SEGMENTATION") OR LIMIT-TO (EXACTKEYWORD , "CHANGE DETECTION")) AND (EXCLUDE (PUBYEAR , 1998) OR EXCLUDE (PUBYEAR , 1993)) AND (LIMIT-TO (LANGUAGE , "ENGLISH")) – 207 results

Title only:

TITLE ((POST-CLASSIFICATION OR "CHANGE DETECTION") AND OBJECT-BASED), Articles post-2000 and written in English – 153 results

Annex 1.2 - Screening Test

To ensure consistency between the primary and secondary reviewers, a screening test was performed using a random sample of 25 articles taken from the search strings. Both the primary reviewer (RB) and secondary reviewer (PR) screened these articles for relevance, based on their titles only. The following criteria were developed and applied for this process:

Crucial terms:

Image Segmentation OR segmentation

Object-based OR Image objects

Post-classification OR change detection (OWTE)

Additional terms:

Accuracy OR Uncertainty OR Error OR Method OR Approach

Review OR Analysis OR Assessment OR Evaluation

The UK OR a location in the UK

Habitat mapping OR Land cover mapping

The use of SAR or VHR imagery

Relevant – Specific mention of 3 or more (including at least two crucial terms):

Probably relevant – 2 crucial terms mentioned OR 1 crucial term and any additional terms

Irrelevant – No crucial terms OR wrong context (urban building/urban cover changes without reference to segmentation approaches/uncertainty analysis OR other topics e.g. brain memory) OR mention use of different classifier to Random Forest

The screening test results showed that the primary reviewer retained 6 additional articles in the relevant folder than the secondary reviewer, who placed them in the irrelevant folder. A discussion of these articles revealed that 2 were mistakes by the secondary reviewer, whilst the remaining 4 were marked as irrelevant because they were too focused on a particular application (such as urban building detection) or focused on a particular element of image analysis less relevant to the review (nadir angles of VHR imagery). Consequently, it was agreed that the primary reviewer would first apply the above screening criteria to article titles, followed by a stricter screening of abstracts in the probably relevant folder, removing articles related to the following:

Niche applications or irrelevant contexts (urban/glacier extents)

Methodological approaches to correcting raw satellite imagery (e.g. misregistration, nadir angles etc)

Comparisons between types of imagery or pixel v object-based approaches

Articles that use imagery significantly outside Sentinel’s range (<5m or >30m)

Image-image change detection, as opposed to map-map change detection

Annex 1.3 - Screening Results

Following the title only screening for each search string, each relevancy folder was combined to get an overall collection of relevant, probably relevant, and irrelevant articles. The probably relevant folder was then screened for abstracts, applying the stricter criteria to assign articles either to the relevant or irrelevant folder. A total of 100 relevant articles were identified, including those which had previously been identified in the initial reading list. Of these, 56 were Open Access articles, which were downloaded into Zotero to be reviewed.

List of abbreviations

ELMS – Environmental Land Management Scheme

EO – Earth Observation

MMU – Minimum Mapping Unit

MFW – Minimum Feature Width

NCEA – Natural Capital and Ecosystem Assessment

OBIA – Object-based Image Analysis

OBCD – Object-based Change Detection

OS – Ordnance Survey

UKCEH – United Kingdom Centre for Ecology and Hydrology

Commission error – An error that occurs when an image classification algorithm assigns a reference data sample to the incorrect habitat class.

Fuzzy clustering – Clustering groups data points into clusters based on their attributes, where data points should be as dissimilar as possible to those in other clusters. Fuzzy clustering is a method whereby data points can be in more than one cluster having been assigned a likelihood of membership to each cluster.

Habitat Class – A habitat category. Example habitat classes include 'Urban', 'Coastal Saltmarsh' and 'Broadleaved Woodland'.

Image Segmentation – The process of clustering pixels in an image into homogenous areas based on the similarity of their spectral, spatial and contextual information. These areas are known as segments or image-objects.

Image Classification – The process of assigning image pixels or objects to a given habitat class. In Living England this is performed using a classification algorithm which is trained to detect each habitat class using a set of reference data.

Machine Learning algorithm – A type of image classification algorithm which learns from its input data to improve and optimise its performance over time.

Omission error - An error that occurs when an image classification algorithm omits a reference data sample from the correct habitat class.

Overall accuracy - The total number of reference data samples that have been assigned by an image classification algorithm to the correct habitat class on a habitat classification map.

Pixel – The smallest item of information in an image. Pixels are arranged in a grid structure, with each pixel representing a square in an image.

Polygon – A spatially explicit shape that represents a given location. In image analysis, polygons are often used as a spatial framework to assign habitat classes to. They can be produced in a number of ways including the image segmentation process (otherwise known as segments or image-objects) or as a result of overlaying objects with inconsistent boundaries (a sliver polygon).

Post-classification Change Detection – An approach to change detection which compares two individual habitat classification maps and assesses changes between them.

Producer's accuracy - Considers the accuracy of a habitat classification map from the map producer's perspective. Understood as the proportion of the reference data samples classified correctly in a given habitat class.

Reference data – Data derived from ground/field data, historical habitat records or desktop surveys of satellite imagery. Used to train the image classification algorithm and validate a habitat classification map.

Region growing – An image segmentation method that generates seed points and then examines neighbouring pixels to decide whether they are similar enough to be included in the same segment as the seed point. This process then iterates to create segments across the image.

Segments – See Image Segmentation

Thresholding – In image processing this is a process for assigning pixels in an image to different categories based on their spectral values. Thresholds are set and categories assigned based on the values being greater than or less than that threshold.

User's accuracy – Considers the accuracy of a habitat classification map from the end-user's perspective. Understood as the proportion of classified data in a habitat class that actually belong to that class.

Catalogue code: NECR460

NATURAL ENGLAND